This is the accepted manuscript (post-print version) of the article. Contentwise, the accepted manuscript version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication (APA)
Please cite the final published version:


Publication metadata

**Title:** A randomized controlled trial of emotion regulation therapy for generalized anxiety disorder with and without co-occurring depression

**Author(s):** Mennin, Douglas S.; Fresco, David M.; O'Toole, Mia Skytte; Heimberg, Richard G.

**Journal:** Journal of Consulting and Clinical Psychology

**DOI/Link:** [http://dx.doi.org/10.1037/ccp0000289](http://dx.doi.org/10.1037/ccp0000289)

**Document version:** Accepted manuscript (post-print)

General Rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.
A Randomized Controlled Trial of Emotion Regulation Therapy
for Generalized Anxiety Disorder with and without Co-Occurring Depression

Douglas S. Mennin
Teachers College, Columbia University

David M. Fresco
Kent State University &
Case Western Reserve University School of Medicine

Mia Skytte O’Toole
Aarhus University

Richard G. Heimberg
Temple University

Correspondence regarding this manuscript can be directed to Douglas S. Mennin, Ph.D., Department of Counseling and Clinical Psychology, Teachers College, Columbia University, 525 West 120th St, New York, NY 10027, USA. Tel: 212-772-5567, mennin@tc.columbia.edu

Funding: This work was supported in part by NIMH [grant number 1 R34 MH70682-01A2] to Richard G. Heimberg. David M. Fresco was supported by National Heart, Lung, and Blood Institute Grant R01HL119977 and National Institute of Nursing Research Grant P30NR015326
Abstract

Objective. Generalized anxiety disorder (GAD) and major depression (MDD), especially when they co-occur, are associated with suboptimal long-term treatment response. One common feature of these disorders is negative self-referential processing (i.e., worry, rumination), which worsens treatment outcome. Emotion Regulation Therapy (ERT) integrates principles from affect science with traditional and contemporary cognitive behavioral treatments to identify and modify the functional nature of negative self-referential processing by targeting motivational and regulatory mechanisms, as well as behavioral consequences. Method. Building on encouraging open trial findings, 53 patients with a primary diagnosis of GAD (43% with comorbid MDD) were randomly assigned to immediate treatment with ERT (n = 28) or a modified attention control condition (MAC, n = 25). Results. ERT patients, as compared to MAC patients, evidenced statistically and clinically meaningful improvement on clinical indicators of GAD and MDD, worry, rumination, comorbid disorder severity, functional impairment, quality of life, as well as hypothesized mechanisms reflecting mindful attentional, metacognitive, and overall emotion regulation, which all demonstrated mediation of primary outcomes. This superiority of ERT exceeded medium effect sizes with most outcomes surpassing conventions for a large effect. Treatment effects were maintained for nine months following the end of acute treatment. Overall, ERT resulted in high rates of high endstate functioning for both GAD and MDD that were maintained into the follow-up period. Conclusions. Findings provide encouraging support for the efficacy and hypothesized mechanisms underlying ERT and point to fruitful directions for improving our understanding and treatment of complex clinical conditions such as GAD with co-occurring MDD.
Public Health Significance:

Generalized anxiety disorder and major depressive disorder are associated with significant role impairment and poor life quality satisfaction. Yet, treatments have not produced sizable gains into the long-term. Emotion Regulation Therapy (ERT), whose framework is drawn from basic and translational affect sciences, demonstrates superior improvement on clinician rated and self-reported anxiety and depression, satisfaction with life, worrying and ruminating, and in self-perceived ability to manage emotions with mindfulness and perspective taking, relative to patients receiving delayed treatment.
A Randomized Controlled Trial of Emotion Regulation Therapy for Generalized Anxiety Disorder with and without Co-Occurring Depression

Generalized anxiety disorder (GAD) is a burdensome condition marked by chronic and excessive worry (American Psychiatric Association, 2013) and considerable suffering. Highly impairing when occurring alone, GAD is especially impacting when it co-occurs with major depressive disorder (MDD; e.g., Whisman, Sheldon, & Goering, 2000; Henning, Turk, Mennin, Fresco, & Heimberg, 2007). Given high rates of comorbidity and overlapping symptoms, GAD and MDD are often regarded as “distress disorders” (e.g., Watson, 2005; also “emotional disorders”; Barlow, Sauer-Zavala, Carl, Bullis, & Ellard, 2014). Despite the availability of efficacious treatments, GAD and MDD frequently exhibit suboptimal long-term treatment response (e.g., Farabaugh et al., 2012; Newman, Przeworski, Fisher, & Borkovec, 2010).

Worry and rumination are core cognitive processes of distress disorders. Although distinguishable from one another in content (i.e., loss versus threat, past versus future, etc.), rumination and worry are psychometrically related (e.g., Segerstrom, Tsao, Alden, & Craske, 2000) and may reflect a core process of negative self-referential processing (NSRP; Northoff, 2007). NSRP occurs in humans especially during discrepancies between a current emotional/motivational state and a representation of the future (i.e., planning), the past (i.e., failures/losses), or an idealized self (i.e., self-criticism) especially given the human capacity for higher-level self-consciousness and the reinforcing properties of this type of cognitive activity for achieving goal states (e.g., Borkovec, Alcaine, & Behar, 2004). This self-conscious ability is crucial for managing a world in which there is ambiguity and uncertainty (e.g., Mennin & Fresco, 2014). Reflecting on one’s past or forecasting one’s future can assist with mental preparation for action toward desired goals and avoid undesired ones. This same ability,
however, can lead to dysfunction.

For instance, NSRP including worry and rumination become negatively reinforced by the perceived reduction in aversive emotions (e.g., Borkovec et al., 2004; Nolen-Hoeksema, Wisco, & Lyubomirsky, 2008) and this aversion may occur most when an emotional state is expected to lead to a contrasting emotional state (i.e., positive emotions may be followed by negative emotions; Newman & Llera, 2011). Further, these processes are associated with considerable deficits in behavioral learning (e.g., Lissek, 2012; Whitmer & Gotlib, 2012), dysfunctional cognitive correlates including intolerance of uncertainty (Dugas et al., 2010) and meta-worry (Wells et al., 2010), emotional non-acceptance and dysregulation (Mennin, Holaway, Fresco, Moore, & Heimberg, 2007; Roemer, Salters, Raffa, & Orsillo, 2005), and interpersonal dysfunction (Przeworski, et al., 2011). NSRP also predicts inferior treatment response and greater relapse (e.g., Jones, Siegle, & Thase, 2008). The poor response produced by CBT for GAD (e.g., Borkovec & Ruscio, 2001) compared to other anxiety and mood conditions may reflect the refractory nature of NSRP and the limitations of traditional CBT to fully target the functional nature of these processes in relation to emotional experience.

Taken together, clinical science clearly implicates NSRP in the course, severity, and treatment of GAD and MDD. However, basic and translational findings from affect science may offer new insights by connecting NSRP to aberrant or excessive neural activity particularly in the anterior and posterior nodes of the default mode network (e.g., Andreeescu et al., 2014; Chen & Etkin, 2013; Hamilton et al., 2011; Whitfield-Gabrieli & Ford, 2012). Consequently, we and others posit that understanding normative and disordered forms of self-referential processing at a neurobehavioral level of analysis holds great promise for advancing clinical care of treatment refractory conditions such as GAD and MDD (e.g., Mennin & Fresco, 2013).
In an effort to integrate and synthesize theory and findings from traditional cognitive behavior therapy and affect science, Mennin and Fresco (2013; 2014) have advanced an emotion dysregulation model that characterizes distress disorders as marked by heightened emotional experience (i.e., motivational intensity) coupled with NSRP, which functions as compensatory strategy to manage the experience of strongly felt emotional and somatic experiences. Building on this framework, Emotion Regulation Therapy (ERT; e.g., Fresco, Mennin, Heimberg, & Ritter, 2013; Mennin & Fresco, 2014) represents a mechanism-targeted intervention that integrates findings from affect science with principles from cognitive behavioral therapy (CBT; see Mennin, Ellard, Fresco, & Gross, 2013). This model posits that dysfunction in distress disorders can best be understood by 1) motivational mechanisms, reflecting the functional and directional properties of an emotional response tendency; (2) regulatory mechanisms, reflecting the alteration of emotional response trajectories utilizing less (i.e., attentional) and more (i.e., metacognitive) cognitively elaborative systems; and (3) contextual learning consequences, reflecting the promotion of broad and flexible behavioral repertoires (Renna, Quintero, Fresco, & Mennin, 2017). With respect to GAD, dysfunction results from a failure in each of these normative systems of functioning. Specifically, GAD is characterized by heightened emotional experience (i.e., motivational intensity) coupled with negative self-referential processing, which functions as compensatory strategy to manage the experience of strongly felt emotional and somatic experiences. Using a motivational framework (i.e., identifying reward- and risk-based impulses), ERT instructs patients to engage in mindful emotion regulation skills to counteract negative self-referential processing (e.g., worry, rumination, and self-criticism) in service of pursuing intrinsically rewarding and goal-directed actions in their lives.
An initial open trial (OT) demonstrated efficacy in reducing symptoms of anxiety and depression as well as decreasing worry, increasing quality of life, and improving social disability from pre- to post-treatment, with gains maintained throughout a follow-up period (Mennin, Fresco, Ritter, & Heimberg, 2015). ERT also resulted in hypothesized mechanism gains, including greater mindful attentional, meta-cognitive, and overall emotion regulation (Mennin et al, 2015) with gains maintained at 3- and 9-month follow-up (Hedge’s gs = .52 to 3.90).

Building on the encouraging findings, the present study sought to further examine the efficacy of ERT using a controlled research design. Consistent with the stage model of psychosocial intervention development (e.g., Onken, Carroll, Shoham, Cuthbert, & Riddle, 2013), the current study utilized a randomized controlled trial (RCT) design comparing patients randomly allocated to immediate ERT with patients assigned to a modified attention control condition (MAC), wherein patients received periodic contact from study therapists to assess clinical status and offer encouragement while they completed assessments and then awaited their turn to receive open label ERT following their participation in the MAC. We hypothesized that ERT would demonstrate superiority to the MAC in reducing symptoms associated with anxiety, depression, and disability, while increasing quality of life. We further hypothesized that ERT would exceed MAC in ameliorating deficits in hypothesized mechanisms reflecting mindful attention and meta-cognitive regulation (i.e., trait mindfulness, decentering, cognitive reappraisal, and trait emotion dysregulation). We also examined long-term outcome (at 3- and 9-month following treatment) of both acutely and post-MAC treated patients to determine the enduring effects of ERT on all study outcomes. Here, we expected the largest change to be found in the acute phase after which trajectories would maintain effects into the follow-up periods.

Importantly, given the deliberate focus in ERT on the development of adaptive regulatory
skills and the successful probing and improvement of these regulation-related variables in our prior OT (Mennin et al., 2015), we also hypothesized that improvements in less and more elaborative regulatory abilities (i.e., trait mindfulness, decentering, cognitive reappraisal, and trait emotion dysregulation) would mediate primary outcomes (i.e., GAD severity, worry, depression, disability, and life satisfaction) of those treated with ERT compared to the MAC condition. Finally, we examined indices of clinical significance in GAD- and MDD-related outcomes to show that improvements reflect gains of considerable magnitude beyond mere statistical change (Kazdin, 1999).

**Method**

*Participants*

Figure 1 displays the CONSORT flowchart for patient enrollment and progression. One Hundred Thirty-Four individuals sought treatment both as part of routine care or specifically in response to directed recruitment efforts for the study (e.g., fliers). All individuals were assessed for general eligibility for clinical care using diagnostic assessment (see below). Fifty-three adults at two psychology training clinic sites in the Northeastern United States (Site 1 was located in XXXXXX, n = 28; Site 2 was located in XXXXXX, n = 25) ultimately expressed interest in the study and consented to be randomized. Institutional Review Boards approved procedures for the study at both sites, and all patients provided informed consent. The targeted sample size was based on a power analysis proposed in the NIMH funded R34 to enroll 34 patients drawing on meta-analytic effect sizes comparing an active treatment to WL control. This proposed sample size also anticipated expected attrition during the RCT of 4 patients leaving 30 completers. As such we would be powered to detect group by time interaction effect sizes ranging from Cohen’s $d = .44$ (no attrition) to 48 (with expected attrition). Seeing as we successfully enrolled 53
patients with 44 completers, our study was actually powered to detect group by time interaction effect sizes ranging from $d = .36$ (full sample) to .38 (completers).

Diagnoses were derived from The Anxiety Disorders Interview Schedule, Lifetime version for DSM-IV (ADIS; site 1) or the Structured Clinical Interview for DSM-IV (SCID; site 2). Interviewers at both sites were clinical psychologists or doctoral students in clinical psychology trained according to the guidelines of the ADIS or SCID. The intake clinician assessed all Axis I disorders and provided a rating of severity using the ADIS clinician severity rating (CSR). A score of 4 or greater (range=0 “no symptom presentation” to 8 “severe symptoms, impairment, or distress”) is given for diagnoses that meet full DSM-IV criteria and are clinically significant.

Patients were required to have a primary or co-primary diagnosis (based on clinical severity) of GAD to be admitted into the study. In addition, an independent assessor also administered the GAD module at pre-, mid-, post-treatment, and at the follow-up assessments. Further, modules with positive diagnoses at pre-treatment were re-administered at post-treatment and into the follow-up periods. Diagnostic agreement (in diagnosis and CSR within 1 point; ADIS CSR was used at both sites) for GAD between the intake clinician and the independent assessor at pre-treatment was necessary for study inclusion. Agreement rate was 100% across both sites.

Additional inclusion criteria consisted of fluency in spoken and written English; and willing and able to give informed written consent. Exclusion criteria consisted of prominent active suicidal ideation/intent; DSM-IV diagnosis of substance abuse or dependence within the previous 6 months; a current DSM-IV diagnosis of organic mental disorder; schizophrenia, psychotic disorder, bipolar I disorder, or dementia; unwillingness to terminate or suspend current psychotherapy; and concurrent psychotropic medication not stabilized for at least 3 months.

The sample was mostly women (75%), with a mean age of 39 ($SD = 14.5$), primarily
White (87%; Asian American/Pacific Islander 2%; African-American 6%; other 6%), non-Hispanic (91%), and well-educated (64% had at least a college education). Twenty-five patients were employed full time followed by working part-time ($n = 7$), being full-time students ($n = 5$), unemployed ($n = 10$), a homemaker ($n = 4$), or retired ($n = 2$). Median annual family income for the sample was $48,819 (Range = $0 to $176,000). Twenty-three patients had a concurrent MDD diagnosis. Thirty-one patients had at least one additional current diagnosis (Range: 1-5) including obsessive compulsive disorder ($n = 7$), panic disorder ($n = 10$), post-traumatic stress disorder ($n = 4$), social phobia ($n = 17$), specific phobia ($n = 7$), dysthymic disorder ($n = 5$), and eating disorder ($n = 1$). Thirteen patients participated while receiving concurrent antidepressant medication; 7 patients enrolled receiving benzodiazepines. Of the 53 consented patients, 28 were randomized to treatment and 25 were randomized to the MAC condition. Group assignment was determined by a random-number generator at the outset of the trial. Patients and clinicians remained unaware of randomization condition until after the signing of informed consent.

Twenty-five patients in the immediate treatment condition (89%) and 20 in the MAC condition (80%) completed the acute phase of the study. Reasons for attrition from acute ERT included a patient’s mother had a major mental health issue that required leaving the area, a patient decided to leave the area to transfer to another university, a patient did not return after reporting an abuse history, a patient began taking psychotropic medication, and divorce proceedings made it difficult for another patient to regularly attend sessions. Twenty MAC completers initiated open-label ERT and 15 completed treatment (75%). Three- and nine-month follow-up assessments were conducted for all patients who completed the treatment. There were no demographic differences across the two sites, and no baseline clinical and demographic differences associated with treatment response (analyses available upon request). Thus, analyses
were conducted on the full sample aggregated across sites. RCT analyses compared ERT to MAC patients at pre-, mid-, and post-acute treatment. Follow-up analyses were conducted on all patients who received ERT (immediate ERT + MAC patients who received open-label ERT).

Treatment

ERT consisted of 20 weekly sessions of 60-minute duration except for Sessions 11-16 which lasted 90 minutes to accommodate exposure exercises. In the first half, sessions focused on teaching emotion regulation strategies via attention regulation and more verbally elaborate meta-cognitive regulation. In the second half of ERT, patients were encouraged to deploy regulation skills during in-session and out-of-session exposures that simultaneously invoked both engagement of a context that is both rewarding and threatening. The final sessions focused on consolidating gains and preparing for termination (Fresco et al., 2013; Mennin & Fresco, 2014).

Ten clinical psychology doctoral students trained to administer ERT by the first and second authors served as protocol therapists, and received 2 hours of weekly group supervision to discuss active cases. All therapists reviewed their session video to generate a detailed timestamped and annotated note detailing how each intervention was received and what issues arose in implementation. Every video and accompanying note was reviewed by the first and second authors prior to supervision. Each case was discussed for approximately twenty minutes with additional time provided when emergent clinical management issues came to light.

Adherence coders were comprised of research assistants or doctoral students trained in coding and taught the components of the therapy by an upper level doctoral student who had served as an ERT therapist for 3 years. Coders read and discussed each ERT session and coded a practice training case to calibrate their ratings. Discrepancies and questions were discussed and resolved during training. Coders assessed the required elements in each session of ERT as
reflected the treatment manual. Coders also assessed adherence 'lapses' (i.e., incorporating non-ERT treatment elements, failing to adhere to the recommended session time) as either absent or present during each session. Coders rated the frequency (0 = therapist did not address component/engage action, 1 = therapist addressed component/engaged action, 2 = therapist addressed component/engaged action in more detail) and skillfulness (0 = therapist was not skillful or performed poorly, 1 = therapist performed action adequately, 2 = therapist performed action very skillfully) of therapist actions consistent with the different ERT components. Any lapses as well as aptitude in performing central ERT components were considered by the rater in their total skillfulness score. Total ratings of frequency and skillfulness of therapist actions were coded across all phases of ERT. This sum was then divided by the overall possible points for frequency and skillfulness. To demonstrate adherence and fidelity to the treatment protocol, 25% of all videos of ERT sessions were coded, with 40% of cases coded by two coders to establish inter-reliability. Overall, the mean skillfulness rating of the therapists coded was 97.8%, while the mean frequency of actions consistent with the treatment protocol was 96.5%. Reliability ratings between the coders were 96.46% for frequency of actions and 95.0% for skillfulness.

*Modified Attention Control (MAC).* Patients assigned to MAC received periodic contact from study therapists to assess clinical status and offer encouragement while they completed assessments and awaited their turn to receive open label ERT following their participation in the MAC. The MAC consisted of four 15-minute telephone contacts occurring at weeks 2, 5, 9, and 13. During these phone contacts, a therapist inquired about the patient’s symptoms and general functioning. If evidence was obtained during any of these telephone contacts that significant clinical deterioration had occurred, the therapist discussed the clinical state of the patient with the PI at each performance site who would make a clinical decision regarding whether the patient
needed to be removed from the study and assisted in obtaining alternative treatment.

Measures

Patients completed clinician-assessed and self-report measures of anxiety, depression, worry, disability, and quality of life as well as hypothesized mechanisms. The assessment schedule for these measures was pre-treatment, mid-treatment, post-acute treatment, and three- and nine-month follow-up. Intake interviewers and independent assessors (all assessment points) were blind to patients’ assignment in the RCT.

Clinician and Diagnostic Assessment.

Independent assessors provided clinical assessment of patients at all points utilizing the ADIS CSR and a modified version of the Clinical Global Impression Rating Scales (CGI; Guy, 1976). A version with anchor points developed specifically for rating improvement (CGI-I) in symptoms associated with GAD was utilized. Patients received a rating of 1 (very much improved) or 2 (much improved) were classified as responders.

Self-report Symptom and Severity Measures

The Penn State Worry Questionnaire (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990) is a widely used 16-item measure of trait worry (α = .85). The 7-item State-Trait Anxiety Inventory (STAI; Bieling, Antony, & Swinson, 1998) is an abbreviated version of this frequently used measure, comprised of the seven items that loaded most highly on an anxiety factor (present sample α = .85). The Mood and Anxiety Symptom Questionnaire-Short Form (MASQ; Watson & Clark, 1991) is a 62-item instrument comprised of four subscales: General Distress Anxious Symptoms (GDA; present sample α = .79), General Distress Depressive Symptoms (GDD; present sample α = .92), Anxious Arousal (AA; present sample α = .89), and Anhedonic Depression (AD; present sample α = .66). The Beck Depression Inventory-II (BDI–II; Beck,
Steer, & Brown, 1996) is a 21-item self-report measure that assesses the affective, cognitive, behavioral, and somatic symptoms of depression (present sample α = .90). The Brooding subscale of *Rumination Scale* (RS; Treynor, Gonzalez, & Nolen-Hoeksema, 2003) is a five-item measure of self-reported rumination uncontaminated by depression symptom content (present sample α = .77; only given at Site 2). The Sheehan Disability Scale (SDS; Sheehan, 1983) is a commonly utilized measure assessing impairment at work, in social relationships, and in responsibilities at home and with family (present sample α = .60). The *Quality of Life Inventory* [QOLI; Frisch, Cornell, Villaneuva, & Retzlaff, 1992] assesses the degree to which an individual is satisfied with 16 areas of his or her life (e.g., health, standard of living, friendships, relationship with family, community, etc.; present sample α = .81).

**Hypothesized Mechanism Measures**

The *Five Facet Mindfulness Questionnaire* (FFMQ; Baer, Smith, Hopkins, Krietemayer, & Toney, 2006) is a 39-item self-report measure assessing trait mindfulness (present sample α = .90). The 11-item Decentering subscale of the *Experiences Questionnaire* (EQ; Fresco et al., 2007) assessed decentering (i.e., the meta-cognitive ability to observe items that arise in the mind with distance and perspective; present sample α = .71). The *Difficulties in Emotion Regulation Scale* (DERS; Gratz & Roemer, 2004) is a 36-item measure of the acceptance of emotions, ability to engage in goal-directed behavior when distressed, impulse control, awareness of emotions, access to strategies for regulation, and clarity of emotions (present sample α = .93). The Cognitive Reappraisal subscale of the *Emotion Regulation Questionnaire* (ERQ; Gross & John, 2003) is a 6-item measure of cognitive reappraisal (present sample α = .91).

The following measures were considered “primary outcomes” given centrality to hypotheses and primary focus of other GAD trials, thus, only these measures were utilized in
mediation analyses: GAD CSR, PSWQ, BDI, SDS, & QOLI. Although MDD CSR and the RSQ are centrally related to hypotheses, there were limitations in data collection of these latter two measures, and, thus, they were not suitable to examine as outcomes in mediation analyses (i.e., MDD CSR was only assessed at pre- and post-treatment; RSQ was only administered at site 2).

**Acute treatment effects**

Mixed linear models (MLMs) were chosen to compare groups over time on primary and secondary outcome variables. MLMs tolerate missing values and thus do not unnecessarily compromise statistical power. All MLMs were based on the intent-to-treat sample \((N = 53)\). Using MLMs, without any ad hoc imputation of missing values, is recommended over other options of handling missing data in longitudinal clinical trials with missing values (Chakraborty & Gu, 2009). The data were hierarchically arranged in two levels, where time at level 1 was nested within individuals at level 2. MLMs of the acute treatment effect included fixed effects for intercept, time (pre-, mid-, and post-treatment), group (ERT or MAC), and a time × group interaction. The models included a random intercept, and the slope was specified as random if it improved the model fit as evaluated by a significant change in the -2LL fit statistics. Dependent variables were treated as continuous and were estimated by the maximum likelihood method. Furthermore, robust standard errors were specified, thereby correcting for potential deviation from a normal distribution and heteroscedasticity.

Effect sizes were calculated based on the residual variance in the models, which represents the recommended approach for MLMs (Selya, Rose, Dierker, Hedeker, & Mermelstein, 2012). Cohen’s \(f\) for the fixed effect \(b\) can be derived from: \((R_{ab}^2 - R_a^2) / (1 - R_{ab}^2)\). We transformed this value into Hedges’ \(g\), thereby adjusting for the small sample size (e.g., Borenstein et al., 2010). Like Cohen’s \(d\), Hedges’ \(g\) is interpreted with conventions of Small =
.20, Medium = .50, Large = .80. Given multiple comparisons and to address the possibility of Type I errors, corrections were performed using the Benjamin-Hochberg procedure with a conservative 0.05 false-discovery rate (McDonald, 2014).

Indirect effects

Mediation analyses followed the principles of lower-level mediation (Krull & MacKinnon, 2001; Kenny, Korchmaros, & Bolger, 2003), where lower-level variables refer to variables measured multiple times (i.e., level 1) for each individual (i.e., level 2). Determination of indirect effects was based on the product-of-coefficients approach. This approach calculates the product of two paths: path a (between the independent variable and the mediator) and path b (between the mediator and the dependent variable controlling for the independent variable). The independent variable here refers to group (ERT vs. MAC), the explored mediators were the FFMQ total score, EQ decentering, ERQ cognitive reappraisal, DERS total score, and the dependent variables were the primary outcome measures. Indirect effects of the proposed mediators were separately investigated.

When having obtained the necessary paths, results were bootstrapped with 5000 iterations in order to obtain both bootstrapped standard errors and bias-corrected confidence intervals (e.g., Krull & MacKinnon, 2001). Bootstrapping is a nonparametric resampling procedure that does not impose the assumption of normality of the sampling distribution. By repeating this process, an empirical approximation of the sampling distribution of $ab$ is built and used to construct confidence intervals for the indirect effect (Preacher & Hayes, 2008).

Follow-up analyses

In the MLMs of the long-term treatment effect, a log-linear function of time was estimated, which demonstrated a superior fit as compared to a linear function of time and
consistent with the hypothesis that the largest change would be found in the acute phase, after which effects would be maintained and, thus, trajectories would be flat in the follow-up period.

Clinical response analyses

As seen in Table 4, clinically significant improvement was assessed in four ways for the combined sample as per Mennin et al. (2015). Simple clinical response was reflected in a CSR score less than 4, which is the established clinical threshold for presence of a diagnosis or a CGI-I score less than 3 (at least “much improved”; only available for GAD). Two more stringent indices of clinical significance were derived from procedures used by Borkovec, Newman, Pinkus, and Lytle (2002) and Ladouceur and colleagues (2000). Specifically, patients were regarded as GAD responders if they demonstrated a clinically meaningful response on at least four of the following six GAD indices: (GAD CSR < 4, CGI-I < 3, at least 30% improvement on the PSWQ, STAI-7, MASQ-AA, and MASQ-GDA). Similarly, patients were regarded as MDD responders if they demonstrated a clinically meaningful response on at least three of four MDD indices (MDD CSR < 4, at least 30% improvement on the BDI-II, MASQ-AD, and MASQ-GDD). Finally, an index of high endstate functioning was derived by assessing whether patients fell into the normative range (within one standard deviation of healthy norms on published clinical measures; Ladouceur et al., 2000) on at least four of six GAD measures (GAD CSR, CGI-I, PSWQ, STAI-7, MASQ-AA, and MASQ-GDA) and three of four MDD measures (MDD CSR, BDI-II, MASQ-AD, and MASQ-GDD). In all four scenarios, these indices of clinical significance yield a ratio where the numerator reflects the number of patients meeting that definition of clinical response divided by the total number of patients receiving ERT.

This approach to assessing clinical significance is consistent with an intention to treat (i.e., ITT) analysis with two important but justifiable modifications (e.g., Abraha et al., 2015).
First, instead of examining solely the cases who received immediate ERT, we wished to examine the clinical impact of ERT on all patients who received treatment. Thus, the denominator reflects the total number of patients receiving ERT in our trial as we elected to aggregate patients randomized to receive ERT ($n=28$) plus patients who received open label ERT following their completion of the MAC ($n=20$). A denominator of 48 was utilized to reflect the total number of patients receiving ERT. Consistent with per protocol analysis, the number of patients counted in the intention to follow (i.e., ITF) analyses was defined as the total number of patients who completed treatment and provided some or all follow-up data ($n=40$). In essence, with ITF, the question we are attempting to address is, “How many patients treated with ERT stay well following treatment?’ Unlike a simple completers analysis, in both ITT and ITF, the respective denominators were held constant to help characterize clinical response.

Finally, to account for missing data in clinical response analyses, a “last observation carried forward” (LOCF) \(^1\) approach was employed by carrying forward the last recorded value.

---

\(^1\) We understand that LOCF is commonly seen as controversial because a key assumption is that participants do not change over time and missing data rarely satisfy definitions for missing at random (e.g., Houck et al., 2004). Often, LOCF underestimates the true treatment effects during the acute treatment phase. In some instances, LOCF overestimates treatment durability where clinical deterioration is highly likely (e.g., Alzheimer’s Disease, European Medicines Agency [EMA], 2011) but is less likely to overestimate effect sizes with respect to emotional disorders such as depression (EMA, 2011). Despite these limitations, even critics concede that LOCF represents a transparent, straight-forward way of handling missing data, and likely reflects a conservative, lower range estimate of true treatment effects (e.g., Houck et al., 2004). Over the course of the peer review process, we investigated other forms of imputation that might possess fewer potential biases but concluded that our data were not well suited to these approaches as our data were not missing at random. In fact, whereas the patients requiring data imputation entered the study with comparable clinical severity to patients without missing data ($F$s $<3.2$, $p$s $>.10$), they were significantly less likely to be clinical responders for GAD and MDD outcomes ($F$s $>5.2$, $p$s $<.03$). The net effect is our use of LOCF resulted in our imputing data from patients with a relatively less strong clinical response with ERT, and if there is bias, it is in the direction of modestly suppressing the potency and durability of ERT findings. Faced with no better imputation approaches, we have opted to retain LOCF so that our results can maintain maximal comparability to other trials.
of a clinical measure into subsequent assessment points. Missing data were minimal for acute treatment ITT analyses (Mid-Treatment, n = 2; Post-Acute Treatment, n = 6). For the ITF analyses, post-acute treatment data became the new baseline for patients entering the no-treatment follow-up period. LOCF was necessary in a small number of instances (three-month follow-up, n = 3; nine-month follow-up, n = 7). Similarly, an important research question we wished to address in this trial was to assess the efficacy of ERT for patients with comorbid GAD+MDD (n = 23). As such, a comparable ITT (n = 23) and ITF (n = 19) was utilized. Analyses were performed in Stata (StataCorp, release 14) and SPSS (IBM Corp, version 22).

**Results**

*Acute treatment effects*

Means and standard deviations for all outcomes at the three acute time points are reported in Table 1. There were no significant between-group differences at pretreatment on any of the measures. The *time x group* interaction effects for all three of the diagnostic measures (GAD CSR and MDD CSR, and mean CSR of additional diagnoses) were significant and of moderate to large effect sizes (Hedge’s gs ranging from .72 to .83). The *time x group* interaction effects for all self-reported anxiety outcomes (PSWQ, STAI-7, MASQ-AA, MASQ-GDA), self-reported depression outcomes (BDI, RSQ Brooding, MASQ-AD, MASQ-GDD), disability/quality of life outcomes (SDS, QOLI), and the hypothesized mechanism variables (FFMQ, EQ Decentering, ERQ Reappraisal, DERS) were significant in the expected direction, ranging in effect size from .51 to 1.50. All results survived the correction for multiple comparisons according to the Benjamini-Hochberg critical value. Result statistics can be found in Table 1.

*Mediation analyses*

Mediation analyses were conducted on the six primary outcomes for the four
hypothesized mechanisms utilized as mediators. Result statistics are reported in Table 2. All proposed mediators showed an indirect effect of group on the primary outcomes.

**Follow-up effects and clinical response**

Means and standard deviations from all patients who received ERT of all outcome variables at the five time points are reported in Table 3. All but one of the effect sizes was of a large magnitude, ranging from .70 to 1.77. Result statistics can be found in Table 3.

The proportion of participants meeting the criteria for treatment response is reported in Table 4. Employing both simple criteria and more stringent indices showed that participants continued to improve over the follow-up period on all the GAD-related response measures. Concerning the MDD-related response measures, the same pattern was detected although there was a slight decrease in response over the follow-up period on the two more stringent indices of clinical response (i.e., 30% improvement and high endstate functioning).

Further, although in the acute treatment phase, no patients initiated medication or additional psychotherapy, at the 3-month follow-up, 1 patient initiated antidepressant medication. At the 9-month follow-up, 2 additional patients initiated antidepressant medication and 1 patient initiated anxiolytic medication. With respect to psychotherapy, 1 patient at the 3-month follow-up and 4 patients at the 9-month follow-up initiated new psychological treatment.

**Discussion**

The current study provides additional evidence for the efficacy of ERT following the promising findings of an earlier open label trial (Mennin et al., 2015). ERT was well tolerated as reflected in the high rate of treatment completion among all patients receiving ERT (83%) as well as among patients with comorbid MDD (83%). In terms of treatment efficacy, patients receiving immediate ERT versus the MAC demonstrated statistically and clinically meaningful improvement on measures of GAD and MDD symptoms, additional anxiety disorder severity,
functional impairment, quality of life, worry, rumination, as well as theoretically consistent mechanism factors reflecting mindful attentional, metacognitive, and overall emotion dysregulation. This superiority of immediate ERT over MAC equaled or exceeded conventions for a medium effect size with many outcomes surpassing conventions for a large effect. Importantly, treatment effects attributable to ERT were largely maintained for nine months following the end of acute treatment.

Complementing these findings, the high-water mark of clinical efficacy within the GAD literature examines whether a treatment can restore high endstate functioning, i.e., whether treatment gains bring patients within one standard deviation of a normative sample on one or more indices. In this present study, GAD patients achieved high endstate functioning if they were restored to normative functioning on at least four of six GAD clinical indicators. A similar index of MDD high endstate functioning (at least 3 of 4 MDD indicators) was also utilized. Overall, ERT was associated with sizable rates of high endstate functioning for both GAD and MDD that were maintained at three and nine months following treatment. Finally, mediation analyses were consistent with the underlying theoretical account of the proposed mechanisms of change (e.g., trait mindfulness, decentering, cognitive reappraisal, trait emotion dysregulation), showing the expected indirect effects of ERT treatment on all primary outcomes including GAD severity, worry, depression symptoms, functional impairment, and quality of life.

One question that arose during the review process was whether the ERT response rates were even across all patients in terms of initial clinical severity. To address this question, we performed a post-hoc analysis examining ERT treatment effects as a function of pre-treatment GAD severity as well as whether or not a patient had a comorbid MDD diagnosis. Specifically, with respect to initial GAD severity, we compared patients receiving an initial CSR of 4 or 5 to
patients with an initial CSR of 6 or 7. Similarly, we compared patients with and without a
diagnosis of MDD (CSR<4 vs CSR ≥4). We conducted subgroup analyses to see whether there
were disproportionate rates of clinical response at post-acute treatment, three-month follow-up,
and nine-month follow-up in GAD and MDD HEF, our most stringent response indicators. Initial
GAD severity was not associated with differential response at any timepoint for GAD HEF ($X^2$'s < 1.10, p’s range from .29 to .71) or MDD HEF ($X^2$'s < 1.10, p’s range from .29 to 1.00).
Similarly, initial comorbid MDD diagnosis was not associated with differential response at any
time point in MDD HEF ($X^2$'s <1.54, p’s range from .21 to 1.00) or GAD HEF ($X^2$’s < 3.7, p’s
range from .09 to .58). These post-hoc analyses indicate that patients across the strata of initial
severity were equally likely to have a positive treatment response with ERT.

The current findings are encouraging, yet must also be interpreted in relation to prior
trials. Traditional CBTs for GAD are efficacious in treating GAD (e.g., Borkovec et al., 2002)
but at the same time, tend to produce effect sizes and response rates that lag behind CBTs
designed to address other anxiety and mood conditions (e.g., Dugas, Anderson, Deschenes, &
Donegan, 2010). For example, the finding that has continued to sound the call to action derives
from the highly cited meta-analysis from Borkovec and Ruscio (2001) reporting that traditional
CBTs, at best, deliver only half their treated patients to high endstate functioning (40-50%). In
contrast, the current trial of ERT utilizing a more conservative definition for clinically significant
change of GAD than these traditional trials found a high endstate functioning of 80% at a 9-
month follow-up period). Similarly, trials utilizing traditional CBT approaches only
demonstrated modest acute treatment effects on indices of depression (Borkovec et al., 2002;
Dugas et al., 2010; Ladouceur et al., 2002) and significant effects were found to return to
pretreatment levels at long-term follow-up (Newman et al., 2010). Further, these earlier trials
did not examine clinical significance of MDD outcomes. In this trial, 63% of the depressed subsample achieved high endstate functioning for MDD into a 9-month follow-up period. Beyond the treatment gains for MDD, ERT also demonstrated reductions in clinician-rated symptom severity for other anxiety disorders that were maintained into the long-term follow-up.

ERT and traditional CBT for GAD share important elements (e.g., Borkovec et al., 2002), including cue detection, skills training, and exposure components. However, there are also notable distinctions. For instance, a primary feature of CBT is the identification and modification of distorted beliefs. This intervention principle emphasizes cognitively elaborative, verbally mediated, components (e.g., cognitive restructuring) to a greater extent than intervention principles that target less elaborative, attentional, components. ERT also shares important elements with mindfulness-based interventions (MBIs) for GAD (e.g., mindfulness-based cognitive therapy; MBCT; Craigie, Rees, Marsh, & Nathan, 2008; Evans et al., 2008, & mindfulness based stress reduction; MBSR; e.g., Hoge et al., 2015), which emphasize the targeting of less elaborative attentional features (e.g., sustained & flexible attention) via guided meditation practices. Although results from these MBI trials are promising, the effects on anxiety outcomes were modest and mood outcomes have largely been poor or minimal given that these trials excluded patients with comorbid mood disorders. Further, these studies have not examined clinical significance or changes in additional comorbid conditions.

Taken together, one possible interpretation of the relatively weaker comparative efficacy of traditional CBT or MBIs relative to ERT is that optimal treatment for conditions such as GAD is achieved by interventions that simultaneously target less cognitively elaborative facets (i.e., attention) and more cognitively elaborative facets (i.e., metacognition) as opposed to focusing treatment largely on one or the other. To more fully substantiate this interpretation, future studies
may be needed to directly compare these approaches while equating treatment duration, intensity, and format (i.e., individual versus group). ERT joins a handful of other contemporary CBTs for GAD (e.g., Barlow et al., 2017; Dugas et al., 2010; Roemer, Orsillo, & Salters-Pedneault, 2008; Wells et al., 2010) that also appear to simultaneously infuse both less and more cognitively elaborative training components within a larger functional framework. Further, these trials have demonstrated comparably high effect size outcomes for GAD, associated anxiety disorders, and MDD as well as improved high endstate functioning levels for GAD on par with the current trial (e.g., Roemer, Orsillo, & Salters-Pedneault, 2008; although this study and others contemporary CBTs noted above did not examine high endstate functioning of MDD).

One challenge to determining how and why these comprehensive CBT packages deliver improved efficacy may lie in elucidating the essential targets of GAD. These approaches introduce a myriad of constructs that are seemingly distinct (e.g., acceptance, mindfulness, values, attention training, emotional awareness) but may functionally be quite related to one other and reflect superordinate categories that well aligned with an affect science-based functional framework. For instance, a goal of the RDoC initiative at NIMH (e.g., Insel et al., 2010) is improving the efficacy of treatments for emotional disorders by identifying common biobehavioral markers such as attention and metacognitive regulation, which have been theorized to underlie transdiagnostic factors common to anxiety and mood pathology (e.g., negative self-referential processing; Mennin & Fresco, 2013; Mennin et al., 2013).

Consistent with this approach, findings in the current trial demonstrated that mindful attention and trait emotion regulation as well as specific metacognitive regulation indices (i.e., decentering, reappraisal) statistically mediated clinical effects on worry, anxiety, depression, disability, and quality of life. These findings are comparable to other studies reporting that
attention regulation and metacognitive regulation markers mediate the relationship of treatment to clinical response in anxiety and depression. In particular, metacognitive regulation including reappraisal (e.g., threat reappraisal, panicogenic beliefs, reappraisal self-efficacy, cost and probability estimates) and self-distancing (e.g., decentering, cognitive defusion) mediated anxiety or depression outcomes in traditional CBT (Goldin et al., 2012; Hofmann et al., 2007; Niles et al., 2014; Smits, Powers, Cho, & Telch, 2004). Trait mindfulness, acceptance, and reductions in experiential avoidance mediated treatment-linked effects of mindfulness- and acceptance-based CBTs (Kearns et al., 2016; Kuyken et al., 2010; Niles et al., 2014; Shahar, Britton, Sbarra, Figueredo, & Bootzin, 2010; van Aalderen et al., 2012; Vøllestad, Sivertsen, & Nielsen, 2011). However, mindfulness-based interventions have also demonstrated mediation by metacognitive variables (i.e., decentering; Hoge et al., 2015; Bieling et al., 2012) and comparative trials with two active treatment arms have shown non-specific effects on both attentional/mindfulness and meta-cognitive variables across both modalities (e.g., Niles et al., 2014). Also, for GAD, other variables such as intolerance of uncertainty and anxiety sensitivity have mediated outcomes for treatments that appear to incorporate both less and more elaborative treatments, as well (e.g., Bomyea et al., 2015). To best parse these different targets, it may be particularly important to increase utilization of more objective assessments of mechanisms in treatment change. To this end, we have begun to examine the role of task-related behavioral (Renna et al., in press) and neural indicators (e.g., Fresco et al., 2017; Raab et al., under review) of attentional and metacognitive regulation in ERT outcome.

A notable finding from this trial was the strong clinical response on depression indicators that were mediated by hypothesized mechanisms. Taken together these findings may indicate that a common factors and/or transdiagnostic approach to treating GAD in concert with potential
MDD characteristics is warranted and lends support to characterization of these disorders as “distress” disorders (e.g., The Unified Protocol; Barlow et al., 2017). Nonetheless, it may also be important to specify the implementation of ERT for different subgroups or pathological processes. It is unclear at this point what individual characteristics may distinguish treatment response. As noted above, baseline severity may not be a strong determinant of differential outcome but other, more independent, factors such as resting state functional neural connectivity commonly implicated with NSRP show promise in specifying treatment outcome in ERT (Fresco et al., 2017). Further, at the process level, worry and rumination have demonstrated both similarity (Segerstrom et al., 2000) and dissimilarity (e.g., Aldao, Mennin, & McLaughlin, 2013). Thus, an important future aim will be to determine whether subtle distinctions in pretreatment levels of these processes warrant further differentiation in intervention components to best address the heterogeneity of presentation in distress disorders such as GAD, MDD, and their co-occurrence.

Despite these promising findings, notable limitations must be considered that may suggest directions for further investigation. First, the sample was relatively small and patients were only followed for nine months. Future studies would ideally treat and follow a larger sample for a longer period of time (~ 2 years) to better determine durability of treatment gains. Second, the sample was largely White and non-Hispanic. Future investigations with more racially and ethnically heterogeneous samples are needed to ideally show that ERT effects generalize to the full population of those suffering with distress disorders. Third, although the MAC included active attention from a clinician during the waitlist period and is appropriate for this stage of treatment development (e.g., Onken et al., 2013), the control comparison was relatively inert and not optimally equated against active treatment such as ERT. Interestingly,
recent RCTs comparing newer efficacious treatments against more established treatments (e.g., Dugas et al., 2010; Hayes-Skelton, Roemer, & Orsillo, 2013) have struggled to demonstrate more than incremental advantages for one treatment over another (but, see Wells et al., 2010). Rather, these findings more clearly reveal the limitations of “horse race” trial methodology to determine essential ingredients for improving treatment of refractory distress disorders such as GAD. The way forward may be in conducting adequately powered mediation studies with multiple assessment points to determine how various targets might be differentially affecting outcome. In recent years, experimental therapeutic intervention designs have flourished and focal, computerized, interventions such as attention bias modification have sought to increase precision but may still lack the comprehensiveness necessary to fully target complex and refractory conditions such as GAD and other distress conditions (Mogg, Waters, Bradley, 2017). Thus, to adequately elucidate target mechanisms of interventions for complicated disorders, trial designs may require a careful blend of traditional multi-arm horse race with experimental therapeutics.

In conclusion, the current findings provide considerable preliminary efficacy for ERT in treating GAD as well as associated MDD and other anxiety problems. Building on these encouraging findings, an important next step will be to determine the optimal duration and intensity of ERT (e.g. Renna et al., in press). On balance, the present results provide encouraging initial support for the efficacy and potential mechanisms of ERT in a sample of patients with GAD with and without comorbid MDD and suggest fruitful directions for improving our understanding and treatment of complicated clinical conditions such as the distress disorders.
References


Hofmann, S. G., Suvak, M. K., Barlow, D. H., Shear, M. K., Meuret, A. E., Rosenfield, D., & Woods,


Roemer L., Salters K., Raffa, S., Orsillo, S.M. (2005). Fear and avoidance of internal experiences in


StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.


### Table 1. Acute treatment effects presented as Means (M), standard deviations (SD), z statistics, and Hedge’s g effect size

<table>
<thead>
<tr>
<th></th>
<th>Emotion Regulation Therapy (n = 28)</th>
<th>Modified Attention Control (n = 25)</th>
<th>time x group interaction effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Mid</td>
<td>Post</td>
</tr>
<tr>
<td><strong>Diagnostic measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAD CSR*</td>
<td>5.6 (0.6)</td>
<td>4.1 (1.1)</td>
<td>3.0 (1.3)</td>
</tr>
<tr>
<td>MDD CSR1</td>
<td>4.3 (0.5)</td>
<td>-</td>
<td>1.4 (2.0)</td>
</tr>
<tr>
<td>Mean CSR of additional diagnoses other than GAD1</td>
<td>4.3 (0.4)</td>
<td>-</td>
<td>1.1 (1.3)</td>
</tr>
<tr>
<td><strong>Self-reported anxiety outcomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSWQ*</td>
<td>67.5 (7.3)</td>
<td>55.3 (10.2)</td>
<td>44.0 (9.9)</td>
</tr>
<tr>
<td>STAI-7</td>
<td>18.6 (4.2)</td>
<td>15.3 (3.1)</td>
<td>12.3 (2.6)</td>
</tr>
<tr>
<td>MASQ AA</td>
<td>29.1 (10.7)</td>
<td>24.0 (6.4)</td>
<td>21.0 (3.8)</td>
</tr>
<tr>
<td>MASQ GDA</td>
<td>30.2 (9.6)</td>
<td>25.1 (5.2)</td>
<td>21.3 (4.7)</td>
</tr>
<tr>
<td><strong>Self-reported depression outcomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDI*</td>
<td>19.8 (9.9)</td>
<td>11.3 (7.7)</td>
<td>5.8 (5.8)</td>
</tr>
<tr>
<td>RSQ Brooding*</td>
<td>12.5 (2.4)</td>
<td>12.9 (3.0)</td>
<td>10.3 (2.3)</td>
</tr>
<tr>
<td>MASQ AD</td>
<td>69.6 (12.3)</td>
<td>61.0 (812.8)</td>
<td>53.0 (114.4)</td>
</tr>
<tr>
<td>MASQ GDD</td>
<td>32.6 (8.8)</td>
<td>27.4 (5.2)</td>
<td>20.5 (5.3)</td>
</tr>
<tr>
<td><strong>Disability/quality of life outcomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDS*</td>
<td>15.1 (6.2)</td>
<td>10.6 (5.7)</td>
<td>7.3 (5.4)</td>
</tr>
<tr>
<td>QOLI*</td>
<td>.7 (1.8)</td>
<td>1.2 (1.5)</td>
<td>2.2 (1.2)</td>
</tr>
<tr>
<td><strong>Proposed mediators</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFMQ total</td>
<td>109.8 (15.5)</td>
<td>122.2 (20.7)</td>
<td>131.6 (17.7)</td>
</tr>
<tr>
<td>ERQ Reappraisal</td>
<td>22.6 (6.2)</td>
<td>26.4 (5.2)</td>
<td>29.7 (4.4)</td>
</tr>
<tr>
<td>EQ Decentering</td>
<td>29.5 (6.7)</td>
<td>33.7 (7.0)</td>
<td>39.2 (5.1)</td>
</tr>
<tr>
<td>DERS total</td>
<td>98.2 (23.1)</td>
<td>82.3 (20.3)</td>
<td>74.6 (19.0)</td>
</tr>
</tbody>
</table>

Note. g = Hedge’s g; GAD=generalized anxiety disorder, CSR=clinician’s severity rating from the Anxiety Disorders Interview Schedule for DSM-IV, Lifetime Version; PSWQ=Penn State Worry Questionnaire; STAI=State Trait Anxiety Inventory; MASQ=Mood and Anxiety Symptom Questionnaire; AA=Anxious Arousal; GDA=General Distress Anxiousness; MDD=Major Depressive Disorder; BDI-II=Beck Depression Inventory-II; RSQ=Response Style Questionnaire; AD=Anhedonic Depression; GDD=General Distress Depression; SDS=Sheehan Disability Scale; QOLI=Quality of Life Inventory; FFMQ=Five Facet Mindfulness Questionnaire; ERQ=Emotion Regulation Questionnaire; EQ=Experiences Questionnaire; DERS=Difficulties in Emotion Regulation Scale; *=primary outcomes; 1=clinician assessment not conducted at Mid-Treatment; 2=only administered at site 2.
### Table 2. Indirect effects of proposed mediators on primary outcomes

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>BSSE</th>
<th>BCLL</th>
<th>BCUL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GAD CSR mediated by:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFMQ total score</td>
<td>-.30</td>
<td>.18</td>
<td>-.73</td>
<td>-.01</td>
</tr>
<tr>
<td>DERS total score</td>
<td>-.39</td>
<td>.18</td>
<td>-.81</td>
<td>-.10</td>
</tr>
<tr>
<td>ERQ Reappraisal</td>
<td>-.27</td>
<td>.13</td>
<td>-.67</td>
<td>-.07</td>
</tr>
<tr>
<td>EQ Decentering</td>
<td>-.41</td>
<td>.19</td>
<td>-.87</td>
<td>-.10</td>
</tr>
<tr>
<td><strong>PSWQ mediated by:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFMQ total score</td>
<td>-2.76</td>
<td>1.36</td>
<td>-6.33</td>
<td>-.40</td>
</tr>
<tr>
<td>DERS total score</td>
<td>-4.32</td>
<td>1.78</td>
<td>-8.51</td>
<td>-1.23</td>
</tr>
<tr>
<td>Reappraisal</td>
<td>-2.79</td>
<td>1.57</td>
<td>-7.40</td>
<td>-1.62</td>
</tr>
<tr>
<td>EQ Decentering</td>
<td>-4.57</td>
<td>1.75</td>
<td>-8.22</td>
<td>-1.36</td>
</tr>
<tr>
<td><strong>BDI mediated by:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFMQ total score</td>
<td>-2.71</td>
<td>1.45</td>
<td>-5.89</td>
<td>-1.12</td>
</tr>
<tr>
<td>DERS total score</td>
<td>-4.02</td>
<td>1.69</td>
<td>-7.65</td>
<td>-1.97</td>
</tr>
<tr>
<td>ERQ Reappraisal</td>
<td>-2.34</td>
<td>1.18</td>
<td>-5.04</td>
<td>-1.34</td>
</tr>
<tr>
<td>EQ Decentering</td>
<td>-3.12</td>
<td>1.26</td>
<td>-5.86</td>
<td>-1.87</td>
</tr>
<tr>
<td><strong>QOLI mediated by:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFMQ total score</td>
<td>.29</td>
<td>.16</td>
<td>.02</td>
<td>.64</td>
</tr>
<tr>
<td>DERS total score</td>
<td>.42</td>
<td>.18</td>
<td>.11</td>
<td>.83</td>
</tr>
<tr>
<td>ERQ Reappraisal</td>
<td>.24</td>
<td>.13</td>
<td>.04</td>
<td>.54</td>
</tr>
<tr>
<td>EQ Decentering</td>
<td>.30</td>
<td>.13</td>
<td>.09</td>
<td>.61</td>
</tr>
<tr>
<td><strong>SDS mediated by:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFMQ total score</td>
<td>-1.58</td>
<td>.86</td>
<td>-3.52</td>
<td>-.10</td>
</tr>
<tr>
<td>DERS total score</td>
<td>-2.40</td>
<td>1.01</td>
<td>-4.51</td>
<td>-.58</td>
</tr>
<tr>
<td>ERQ Reappraisal</td>
<td>-1.48</td>
<td>.74</td>
<td>-3.15</td>
<td>-.21</td>
</tr>
<tr>
<td>EQ Decentering</td>
<td>-1.61</td>
<td>.69</td>
<td>-3.17</td>
<td>-.47</td>
</tr>
</tbody>
</table>

*Note. BSSE=Bootstrapped standard error, BCLL=Bias-corrected lower level of 95% confidence interval, BCUL=Bias-corrected upper level of 95% confidence interval, GAD=generalized anxiety disorder, CSR=Clinician’s Severity Rating from the Anxiety Disorders Interview Schedule for DSM-IV, Lifetime Version; PSWQ=Penn State Worry Questionnaire; BDI-II=Beck Depression Inventory-II; RSQ=Response Style Questionnaire; SDS=Sheehan Disability Scale; QOLI=Quality of Life Inventory; FFMQ=Five Facet Mindfulness Questionnaire; ERQ=Emotion Regulation Questionnaire; EQ=Experiences Questionnaire; DERS=Difficulties in Emotion Regulation Scale. Results refer to the indirect effect of the proposed mediators investigated separately. Significant (p<.05) indirect effects are in bold.*
Table 3. Long-term treatment effects presented as Means (M), standard deviations (SD), z statistics, and Hedge’s g effect size

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Time point</th>
<th>Effect of time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Mid</td>
</tr>
<tr>
<td>Diagnostic measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*GAD CSR</td>
<td>5.4 (0.9)</td>
<td>4.3 (1.2)</td>
</tr>
<tr>
<td>MDD CSR¹</td>
<td>4.5 (0.5)</td>
<td>2.8 (1.8)</td>
</tr>
<tr>
<td>Mean CSR of additional diagnoses other than</td>
<td>4.4 (0.6)</td>
<td>2.7 (1.4)</td>
</tr>
<tr>
<td>GAD³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-reported anxiety</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*PSWQ</td>
<td>67.9 (8.7)</td>
<td>57.3 (10.2)</td>
</tr>
<tr>
<td>STAI-7</td>
<td>19.1 (4.2)</td>
<td>15.9 (3.3)</td>
</tr>
<tr>
<td>MASQ AA</td>
<td>30.0 (6.9)</td>
<td>26.0 (8.4)</td>
</tr>
<tr>
<td>MASQ GDA</td>
<td>29.8 (6.9)</td>
<td>25.5 (6.0)</td>
</tr>
<tr>
<td>Self-reported depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*BDI</td>
<td>19.2 (10.5)</td>
<td>11.7 (8.8)</td>
</tr>
<tr>
<td>*RSQ Brooding ±</td>
<td>13.2 (2.9)</td>
<td>13.4 (2.5)</td>
</tr>
<tr>
<td>MASQ AD</td>
<td>71.1 (13.9)</td>
<td>63.7 (13.8)</td>
</tr>
<tr>
<td>MASQ GDD</td>
<td>33.2 (10.5)</td>
<td>28.5 (9.6)</td>
</tr>
<tr>
<td>Disability/quality of life</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*SDS</td>
<td>15.1 (6.6)</td>
<td>12.6 (6.3)</td>
</tr>
<tr>
<td>*QOLI</td>
<td>0.2 (1.8)</td>
<td>0.8 (1.7)</td>
</tr>
<tr>
<td>Proposed mediators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFMQ total</td>
<td>111.2 (19.1)</td>
<td>118.0 (18.9)</td>
</tr>
<tr>
<td>ERQ Reappraisal</td>
<td>23.1 (8.1)</td>
<td>26.4 (6.1)</td>
</tr>
<tr>
<td>EQ Decentering</td>
<td>29.9 (6.6)</td>
<td>32.9 (6.0)</td>
</tr>
<tr>
<td>DERS total</td>
<td>97.0 (22.6)</td>
<td>91.2 (20.5)</td>
</tr>
</tbody>
</table>
Note. $g =$ Hedge's $g$ effect size; GAD=generalized anxiety disorder, CSR=Clinician’s Severity Rating from the Anxiety Disorders Interview Schedule for DSM-IV, Lifetime Version; PSWQ=Penn State Worry Questionnaire; STAI=State Trait Anxiety Inventory; MASQ=Mood and Anxiety Symptom Questionnaire; AA=Anxious Arousal; GDA=General Distress Anxiousness; MDD=Major Depressive Disorder; BDI-II=Beck Depression Inventory-II; RSQ=Response Style Questionnaire; AD=Anhedonic Depression; GDD=General Distress Depression; SDS=Sheehan Disability Scale; QOLI=Quality of Life Inventory; FFMQ=Five Facet Mindfulness Questionnaire; ERQ=Emotion Regulation Questionnaire; EQ=Experiences Questionnaire; DERS=Difficulties in Emotion Regulation Scale; *=primary outcomes; 1=Clinician assessment not conducted at Mid-Treatment; 2=Only administered at site 2.
### Table 4. Percentages of patients receiving ERT meeting criteria for treatment response

<table>
<thead>
<tr>
<th></th>
<th>Post-treatment:</th>
<th>3-month follow-up:</th>
<th>9-month follow-up:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># responders</td>
<td># responders</td>
<td># responders</td>
</tr>
<tr>
<td></td>
<td>(%) ITT</td>
<td>(%) ITF</td>
<td>(%) ITF</td>
</tr>
<tr>
<td><strong>GAD clinical response</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADIS GAD CSR &lt; 4</td>
<td>33/48 (69%)</td>
<td>29/40 (73%)</td>
<td>35/40 (88%)</td>
</tr>
<tr>
<td>CGI-improvement &lt; 3</td>
<td>37/48 (77%)</td>
<td>36/40 (90%)</td>
<td>39/40 (98%)</td>
</tr>
<tr>
<td>30% improvement (4+ of 6 criteria met)</td>
<td>36/48 (75%)</td>
<td>35/40 (88%)</td>
<td>36/40 (90%)</td>
</tr>
<tr>
<td>High endstate functioning (4+ of 6 criteria met)</td>
<td>32/48 (67%)</td>
<td>30/40 (75%)</td>
<td>32/40 (80%)</td>
</tr>
<tr>
<td><strong>MDD clinical response</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADIS MDD CSR &lt; 4</td>
<td>15/23 (65%)</td>
<td>15/19 (79%)</td>
<td>16/19 (84%)</td>
</tr>
<tr>
<td>CGI-improvement &lt; 3</td>
<td>16/23 (70%)</td>
<td>15/19 (79%)</td>
<td>17/19 (89%)</td>
</tr>
<tr>
<td>30% improvement (3+ of 4 criteria met)</td>
<td>13/23 (57%)</td>
<td>14/19 (74%)</td>
<td>12/19 (63%)</td>
</tr>
<tr>
<td>High endstate functioning (3+ of 4 criteria met)</td>
<td>14/23 (61%)</td>
<td>14/19 (74%)</td>
<td>12/19 (63%)</td>
</tr>
</tbody>
</table>

**Note.** ITT = Intention to Treat; ITF = Intention to Follow; GAD=generalized anxiety disorder; CSR=Clinician’s Severity Rating from the Anxiety Disorders Interview Schedule for DSM-IV, Lifetime Version; CGI=clinical global impression; MDD=major depressive disorder.
Figure Captions

**Figure 1.** CONSORT diagram.