Emotion-focused therapy for social anxiety disorder: Results from a multiple-baseline study

Article in Journal of Consulting and Clinical Psychology · March 2017
DOI: 10.1037/ccp0000166

CITATIONS
19

READS
1,187

3 authors, including:

Ben Shahar
Hebrew University of Jerusalem
27 PUBLICATIONS 500 CITATIONS

Eran Bar-Kalifa
Ben-Gurion University of the Negev
51 PUBLICATIONS 227 CITATIONS

Some of the authors of this publication are also working on these related projects:

Project
Personality and health View project
Emotion-Focused Therapy for Social Anxiety Disorder: Results From a Multiple-Baseline Study

Ben Shahar  
Interdisciplinary Center Herzliya, Israel

Eran Bar-Kalifa  
Bar-Ilan University

Eve Alon  
Interdisciplinary Center Herzliya, Israel

Objective: The purpose of the present study was to evaluate the efficacy of emotion-focused therapy (EFT) for adults suffering from social anxiety disorder (SAD). Method: Using a nonconcurrent multiple-baseline design, 12 patients (mean age = 26.75 years, SD = 5.15; 7 males) meeting criteria for SAD were treated with up to 28 sessions of EFT. EFT was based on an empathic relationship, 2-chair work for self-criticism, empty-chair work for unresolved feelings, and focusing. Patients were randomized to wait 4, 8, or 12 weeks between the intake and the first therapy session. Intake assessment included the MINI International Neuropsychiatric Interview (MINI; Sheehan et al. 1998), the clinician-administered Liebowitz Social Anxiety Scale (LSAS; Liebowitz, 1987), and various self-report questionnaires. The LSAS was also administered at the end of the baseline period and at posttreatment. The MINI was administered again at posttreatment. Self-reports were administered throughout the baseline, before each therapy session, and at 6-month and 12-month follow-ups. Results: One patient dropped out prematurely. Of the 11 completers, 7 did not meet criteria for SAD at the end of treatment. Intent-to-treat analysis showed that LSAS scores did not change during baseline, significantly improved during treatment (Cohen’s $d = 2.37$), and remained improved during follow-up. Mixed regression models showed that SAD symptoms and self-criticism did not change during baseline, significantly improved during treatment, and remained improved during follow-up. Self-reassurance improved significantly during the follow-up phase. Conclusion: This study provides initial evidence supporting the efficacy of EFT for SAD.

Social anxiety disorder (SAD) is one of the most common and chronic psychological disorders, with a lifetime prevalence of 12.1% in the United States (Kessler, Stein, & Berglund, 2005; Ruscio et al., 2008). Individuals struggling with SAD experience intense fear and shame in social and performance situations related to the possibility of being exposed as flawed and negatively judged. They are also anxious about future social and performance situations and tend to avoid them (American Psychiatric Association, 2013). The result is often severe functional impairment (Kessler, 2003; Kessler, Stein, & Berglund, 1998), isolation (Teo, Lerrigo, & Rogers, 2013), and additional psychopathological comorbidity (Fehm, Pelissoio, Furmark, & Wittchen, 2005).

Cognitive and behavioral therapies are the most studied treatments for SAD, and several meta-analytic studies have shown that these therapies produce moderate to large effect sizes (Acarturk, Cuijpers, van Straten, & de Graaf, 2009; Mayo-Wilson et al., 2014). However, although cognitive–behavioral therapies produce encouraging results, and continue to evolve and improve (Clark et al., 2006), several researchers have emphasized that there are still many patients who do not respond well to CBT or who remain considerably symptomatic at the end of treatment (Davidson et al., 2004; Hofmann & Bögels, 2006; Moscovitch, 2009). As a result, in recent years, there have been a few attempts to adapt and evaluate other treatment approaches to SAD, such as interpersonal (Stangier, Schramm, Heidenreich, 2010).
Emotion-focused therapy (EFT; Greenberg, 2011) is a form of experiential therapy, combining person-centered relational principles (Rogers, 1951) with more directive evocative interventions from Gestalt and experiential therapies. EFT aims at helping patients attend to their emotions, accept them, symbolize them in words, make sense of them, regulate them, and, ultimately, to restructure maladaptive emotion schemes that underlie patients’ symptoms (Greenberg, Rice, & Elliott, 1993; Greenberg & Watson, 2006).

Several research findings suggest that EFT may be a compelling treatment option for SAD. First, recent studies have shown that individuals with SAD experience several emotion-related difficulties, such as experiential avoidance (Kashdan et al., 2014) and poor understanding of their own and others’ emotions (O’Toole, Hoagard, & Mennin, 2013). EFT is fundamentally based on an empathic exploratory stance that helps patients attend to their feelings, create new meanings from them, and integrate daily emotional responses into their life narrative, a process that increases their emotional awareness. EFT therapists, by being empathically attuned, moment by moment, to patients’ emotional states, help patients label their emotions, a process that has recently been shown to be effective in reducing public speaking anxiety (Niles, Craske, Lieberman, & Hur, 2015). Second, socially anxious individuals are highly self-critical (Cox, Fleet, & Stein, 2004; Cox et al., 2000). A central intervention in EFT, the gestalt-derived two-chair dialogue task, is directly designed to alleviate self-criticism by enacting the self-critical attacks and evoking the resulting feelings. During this dialogue, self-criticism is often diminished by the activation of empowering anger, grieving, and self-compassion (Shahar et al., 2012). Third, early histories of emotional abuse, neglect, and other problematic parental behaviors, such as lack of warmth and overprotectiveness, have been consistently shown to play an important role in the development of SAD (Arrindell et al., 1989; Bandelow et al., 2004; Bruce, Heimberg, Blanco, Schneier, & Liebowitz, 2012; Knappe, Beesdo-Baum, Fehm, Lieb, & Wittchen, 2012; Kuo, Goldin, Werner, Heimberg, & Gross, 2011; Simon et al., 2009). These experiences are associated with the formation of maladaptive emotion schemes (often shame-based) that underlie the symptomatic anxiety. A central intervention in EFT, the empty-chair dialogue task for unresolved feelings (Greenberg & Malcolm, 2002; Paivio & Greenberg, 1995), helps patients access, process, and restructure these maladaptive shame-based schemes using an imaginary dialogue with the abusing figure (usually an attachment figure, but can also be bullying peers). Finally, SAD is strongly comorbid with depression, and EFT has been shown to be efficacious with depressed patients (Ellison, Greenberg, Goldman, & Angus, 2009; Goldman, Greenberg, & Angus, 2006; Greenberg & Watson, 1998; Watson, Gordon, Stermac, Kalogerakos, & Steckley, 2003). Like SAD, depression is associated with high levels of shame (Gilbert, 2000; Kim, Thibodeau, & Jorgensen, 2011). Therefore, it is reasonable to extrapolate that EFT interventions that have been shown to ameliorate depressive symptoms may also be efficacious with individuals suffering from SAD.

EFT conceptualizes SAD as a shame-based disorder whereby individuals experience themselves as flawed, worthless, and inferior as a result of an interaction between genetic and temperamental predispositions with early experiences of shaming and rejection by attachment figures or peers (Shahar, 2014). These early emotional injuries lead to an internalized and lasting sense of self as defective and vulnerable to negative judgment and social exclusion (Matos & Pinto-Gouveia, 2010; Pinto-Gouveia, & Matos, 2011). According to EFT theory, this shame-based scheme underlies the symptomatic (secondary, in EFT terms) anxiety. Individuals with SAD are afraid that flaws might be exposed and cause further activation of shame-based feelings. Therefore, they use a variety of avoidance strategies to prevent flaw exposure. One can say that they are anxious about experiencing shame. Self-criticism, for example, serves as an avoidant function because it prevents social risk taking that might result in flaw exposure and shame activation (Gilbert & Procter, 2006; Shahar, Doron, & Szepsenwol, 2015). The primary therapeutic goal in EFT is to access and activate shame during therapy sessions in order to help patients tolerate and regulate it, and also to make it amenable to transformation by exposing it to previously avoided adaptive anger, sadness, and self-compassion (Shahar, 2014).

The Present Study

The purpose of the current study was to assess the efficacy of EFT for SAD. We used a nonconcurrent multiple-baseline single-case design (also called staggered baseline design) in which 12 patients meeting criteria for SAD waited different lengths of time between their initial intake interview and the first therapy session. The term nonconcurrent is used because the baseline and therapy phases of the study occurred at different points in time for each patient (i.e., all patients did not start the baseline period at the same time; Carr, 2005). As common in clinical settings, their intake interview occurred as they were recruited. There are several advantages of using the multiple-baseline design: Because this is one of the first attempts to examine the efficacy of EFT with socially anxious patients (see also Elliott, 2013), it seemed that using a randomized clinical trial and comparing EFT with other active treatments would be premature. Furthermore, multiple-baseline design is superior to an open trial because it controls for the passage of time and also establishes a cause-and-effect association between treatment and outcomes, as participants serve as their own controls (Barlow & Nock, 2009). Our specific hypotheses were that most patients would not meet criteria for SAD at the end of treatment, and that social anxiety symptoms and self-criticism would not change during the baseline period (regardless of its length), would significantly improve during the therapy phase, and would remain improved during the 1-year follow-up phase. We also hypothesized that self-reassurance (a construct akin to self-compassion; see Lange et al., 2010) would not change during the baseline period, would significantly increase during the therapy phase, and would remain improved during the follow-up phase.
Method

Participants

Participants were 12 treatment-seeking patients suffering from SAD. Inclusion criteria were (a) primary diagnosis of SAD as assessed with the Mini International Neuropsychiatric Interview (Sheehan et al., 1998); (b) a score of >28 on the Social Phobia Inventory (SPIN; Connor et al., 2000); (c) Age 18 to 65 years; (d) fluency in Hebrew or English; and (e) willingness to participate in the study and to sign a consent form. Exclusion criteria were (a) imminent suicide risk, self-harm, or other factors suggesting that immediate treatment is needed and waiting would not be advised; (b) current substance dependence; (c) past or current psychotic disorders; (d) past or current bipolar disorder; and (e) currently receiving another form of psychological treatment. Use of psychopharmacology was allowed if already stable (over 3 months) at the time of recruitment.

Procedure

Following institutional review board approval, participants were recruited through posted announcements and e-mails sent to various listservs. Interested participants were first asked to complete the SPIN online to ascertain whether they scored above 28. Eligible participants were phone screened to provide initial information about the structure of the study and to provide an initial, informal assessment of SAD symptoms. Those who agreed to participate after knowing more about the study structure and who indicated anxiety in, and avoidance of, social situations were invited to a structured intake session that included the MINI; the clinician-administered Liebowitz Social Anxiety Scale (LSAS; Liebowitz, 1987); other self-report questionnaires assessing SAD symptoms, depressive symptoms, and self-criticism (described in the Measures section); and an informed consent process. All participants signed the consent form.

Participants passing the structured intake session were then randomized to wait 4, 8, or 12 weeks before their first therapy session (four participants per length). Randomization was conducted by an independent evaluator using the RANDBETWEEN function in Microsoft Excel. Participants were randomized as they were recruited, such that the first participant could be allocated to each of the three baseline lengths and the 12th participant was allocated to the only group that remained open. During the baseline period, participants completed the self-report questionnaires once every 2 weeks. After the baseline period and immediately before the first therapy session, they met with the independent evaluator again to complete the clinician-administered LSAS. During the therapy phase, they completed the self-report questionnaires before each session. Immediately after the treatment phase, they met with the evaluator again to complete the MINI and the clinician-administered LSAS. During the follow-up phase, they completed the self-report questionnaires at 6-month and 12-month posttreatment. All self-report questionnaires were completed online. During the baseline and the follow-up phases, participants were asked to complete the questionnaires from home. During the therapy phase, participants completed the questionnaires before each session using a computer in a room next to the therapy room. Table 1 presents the study’s assessment schedule.

Measures

The MINI International Neuropsychiatric Interview (Sheehan et al., 1998). The MINI is a relatively short, but psychometrically sound, structured interview used to assess 17 common Axis I disorders. It was used to ascertain SAD diagnosis, assess comorbid conditions, and to identify conditions that merited exclusion from this study. The MINI was administered by a doctoral clinical psychology student with extensive training in administering this instrument.

The Liebowitz Social Anxiety Scale (Liebowitz, 1987). We used the clinician-administered LSAS at the intake session, at the end of the waiting period (right before the first therapy session), and at posttreatment. In addition, we used the self-report version of the LSAS at the 6-month and 12-month follow-up evaluations (these evaluations were completed solely online). Both the clinician-administered and the self-report versions of the LSAS are widely used and have excellent psychometric properties (Fresco et al., 2001; Heimberg et al., 1999). Furthermore, several studies have found that the self-report version compares very well with the clinician-administered version (Baker, Heinrichs, Kim, & Hofmann, 2002; Fresco et al., 2001). The LSAS assesses the degree of anxiety and avoidance in 24 social and performance situations. In the self-report version, respondents are asked to indicate the extent to which they are likely to experience fear or avoid specific social or performance situations on a 4-point Likert scale. In the clinician-administered
version of the LSAS, the patient is still provided with the scale anchors and provides his or her rating for each social situation, but the clinician can probe for more information. Although the LSAS can produce several subscales, in this study, we used total scores computed by summing both fear and avoidance ratings across all 24 social situations. Scores could range from 0 to 144, with higher scores reflecting more severe social anxiety.

**Social Phobia Inventory (Connor et al., 2000).** The SPIN is a 17-item widely used and well-validated instrument assessing core symptoms of social anxiety, such as fear and avoidance in social situations, physiological symptoms of anxiety, fear of embarrassment and authority, and fear of criticism and negative judgment. Scores on the SPIN are summed and can range from 0 to 68, with higher scores reflecting more anxiety symptoms.

**Brief Fear of Negative Evaluation Scale (BFNE; Leary, 1983).** The BFNE is also a widely used and well-validated instrument assessing self-criticism and a sense of inadequacy at times of failure and setbacks (“There is a part of me that puts me down”). The BFNE contains three subscales: (a) Inadequate-Self, which captures a moderate level of self-criticism; (b) Hated-Self, which captures a more destructive, disgust-based form of self-criticism (“I am so angry with myself that I want to hurt or injure myself”); and (c) Reassuring-Self, which captures the tendency to be self-supportive and self-reassuring in times of stress (“I am gentle and supportive with myself”). The BFNE has excellent psychometric properties (Gilbert et al., 2004) and its factor structure was recently replicated in clinical and nonclinical samples (Castilho, Pinto-Gouveia, & Duarte, 2015).

**Beck Depression Inventory-II (BDI-II; Beck, Steer, Ball, & Ranieri, 1996).** The BDI is a 21-item inventory assessing symptoms of depression. Each item is scored on a scale of 0 to 3, and total scores can range from 0 to 63, with higher scores indicating more symptoms. The BDI-II is one of the most widely used inventories for depression and it has excellent psychometric properties (Beck et al., 1996).

**Truax Accurate Empathy Scale (Truax, 1967).** The Truax Accurate Empathy Scale is a 9-point observer-rated scale designed for the assessment of empathic accuracy.

---

**Table 1**

<table>
<thead>
<tr>
<th>ID</th>
<th>Age</th>
<th>Gender</th>
<th>Comorbidity</th>
<th>Psychotropic medications</th>
<th>Education</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35</td>
<td>F</td>
<td>GAD, BN</td>
<td>None</td>
<td>MA degree</td>
<td>Single</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>M</td>
<td>GAD</td>
<td>None</td>
<td>MA student</td>
<td>Single</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>M</td>
<td>GAD</td>
<td>ADM</td>
<td>BA degree</td>
<td>In a serious relationship</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>M</td>
<td>None</td>
<td>None</td>
<td>BA student</td>
<td>Single</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>M</td>
<td>None</td>
<td>None</td>
<td>MA degree</td>
<td>In a serious relationship</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>M</td>
<td>PD, MDD</td>
<td>None</td>
<td>BA student</td>
<td>Single</td>
</tr>
<tr>
<td>7</td>
<td>26</td>
<td>M</td>
<td>PD, MDD</td>
<td>ADM/beta-blockers</td>
<td>BA student</td>
<td>In a serious relationship</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
<td>F</td>
<td>PD, GAD</td>
<td>ADM</td>
<td>BA student</td>
<td>Single</td>
</tr>
<tr>
<td>9</td>
<td>26</td>
<td>F</td>
<td>PD, MDD</td>
<td>ADM/benzodiazepines</td>
<td>MA student</td>
<td>Single</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>M</td>
<td>None</td>
<td>None</td>
<td>BA student</td>
<td>Single</td>
</tr>
<tr>
<td>11</td>
<td>34</td>
<td>F</td>
<td>History of AN</td>
<td>None</td>
<td>MA degree</td>
<td>Married</td>
</tr>
<tr>
<td>12</td>
<td>22</td>
<td>F</td>
<td>None</td>
<td>None</td>
<td>BA student</td>
<td>Married</td>
</tr>
</tbody>
</table>

Note. ID = Identification; F = female; M = male; GAD = generalized anxiety disorder; BN = bulimia nervosa; PD = panic disorder; MDD = major depressive disorder; AN = anorexia nervosa; MA = master’s; BA = bachelor; ADM = antidepressant medication.

*Participant #10 discontinued treatment prematurely after 10 sessions.

**Table 2**

**The Study’s Assessment Schedule**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Initial online screening</th>
<th>Intake (first baseline assessment)</th>
<th>Biweekly baseline assessments</th>
<th>End of baseline</th>
<th>Presession assessments</th>
<th>Posttreatment assessment</th>
<th>Follow-up assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPIN</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MINI</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>LSAS-CA</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>LSAS-SR</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>BFNE</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FSCRS</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>BDI-II</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Note. SPIN = Social Phobia Inventory; MINI = The MINI International Neuropsychiatric Interview; LSAS-CA = Liebowitz Social Anxiety Scale – Clinician-Administered; LSAS-SR = Liebowitz Social Anxiety Scale – Self-Report; BFNE = Brief Fear of Negative Evaluation Scale; FSCRS = Forms of Self-Criticizing/Self-Reassuring Scale; BDI-II = Beck Depression Inventory-II.
to measure therapist empathy. Low scores indicate being nonresponsive, giving advice, or detracting from patients’ affective and meaning content. A score of 5 indicates responses that are interchangeable with patients’ responses, and higher scores indicate that therapists’ responses add additional, deeper affective meaning to what the patient has said. This measure has shown good inter-rater reliability (range = .73 to .86) and predictive validity in client-centered therapy (Kiesler, 1973).

**Task-Specific Intervention Adherence Measures (Greenberg & Watson, 1998).** The Task-Specific Intervention Adherence Measures are two 7-point adherence measures—one for two-chair work for self-criticism and one for empty-chair work for unresolved feelings. Each of these tasks is based on a series of therapist interventions, ranging from initiating the task to resolution. Therefore, each point on these scales represents a therapist action, with higher numbers representing interventions closer to resolution. For example, in the two-chair scale, a score of 1 represents initiating the task and facilitating an engaged dialogue between the inner critic and the experiencing self, a score of 4 represents the experiencing-self expressing anger and sadness and expressing needs, and a score of 6 represents softening of the inner critic. In the empty-chair scale, a score of 1 indicates evoking the image of the other in the empty chair, a score of 4 indicates expression of primary emotions and needs, and a score of 7 indicates developing empathy and forgiveness toward the other. All therapist talk-turns in a given segment are coded as adhering or not to one of the steps in the task model. These measures have demonstrated interrater reliability (intraclass correlation coefficient [ICC] range from .76 to .89) and have been found to reliably discriminate EFT from client-centered therapy (Greenberg & Watson, 1998).

**Treatment and Therapists**

EFT (Greenberg, 2011; Greenberg et al., 1993) is a brief, empirically supported treatment for depression (American Psychological Association Division 12 Society of Clinical Psychology, 2013; Ellison et al., 2009; Goldman et al., 2006; Greenberg & Watson, 1998; Watson et al., 2003) that has also shown promising effects in treating symptoms resulting from severe and repeated childhood abuse (Paivio, Jarry, Chagigiorgis, Hall, & Ralston, 2010; Paivio & Nieuwenhuis, 2001). EFT is based on combining client-centered relational elements (unconditional positive regard, congruence, and empathy) with marker-guided experiential interventions designed to facilitate and deepen emotional processing. After creating a strong bond with patients, EFT therapists identify particular markers that call for specific interventions. For example, when patients are self-critical, therapists suggest a two-chair dialogue intervention, and when patients show lingering unresolved feelings toward attachment figures, therapists often suggest an empty-chair dialogue. When patients have a vague or unclear sense of what they are feeling, this calls for focusing (Gendlin, 1996). Length of therapy in this study was 24 to 28 weekly sessions ($M = 26.54, SD = 1.50$). The two therapists (the first and third authors) were psychologists with doctoral degrees in clinical psychology and extensive training in EFT. Each therapist treated six patients.

**Treatment Adherence**

During the therapy phase of this study, therapists met with an EFT expert supervisor once a week for two hours. During these supervision meetings video-taped sessions were reviewed to ensure proper implementation of EFT. In addition, adherence for empathic attunement and adherence to two central EFT task interventions (two-chair intervention for self-criticism and empty-chair dialogue for unresolved feelings) were checked after the therapy phase was completed. Ratings were conducted by an expert EFT therapist and a clinical psychology doctoral student with extensive training in EFT. Because both raters were English speakers, only tapes from English-speaking patients were reviewed (six patients).

For empathy adherence, one session containing no chair-work was selected from Sessions 1 to 3. The middle 20 min of that session was further divided into 4-min video clips, and all 30 clips were pooled together and the two raters rated each segment independently using the Truax Accurate Empathy Scale. For adherence to the EFT task interventions, one session containing two-chair work for self-criticism and one session containing empty-chair work for unresolved feelings was each randomly selected from its corresponding pool. In each session, the entire chair-work segment was rated using the Task-Specific Intervention Adherence Measures.

**Data Analysis**

A one-way analysis of variance was used to assess changes in LSAS scores. A two-level multilevel modeling (assessments nested within patients) was used to assess differences between the baseline, treatment, and follow-up phases, as well as the linear change within each phase for the SPIN, the BFNE, and the three subscales of FSCR. For each of the five outcomes, the fixed model part consisted of (a) an intercept; (b) a dummy indicator variable for the treatment phase; (c) a dummy indicator variable for the follow-up phase; (d) a time-within-baseline variable that was set to equal zero for assessments outside the baseline phase, whereas inside the baseline phase was centered around the first baseline assessment; (e) a time-within-treatment variable that was set to equal zero for assessments outside the treatment phase, whereas inside the treatment phase was centered around the last session; and (f) a time-within-follow-up variable that was set to equal zero for assessments outside the follow-up phase, whereas inside the follow-up phase was centered around the last follow-up assessment. Importantly, the inclusion of the time-within-baseline, time-within-treatment, and time-within-follow-up variables in the model (with this particular centering) turned (a) the intercept as an indicator of the outcome at the first baseline assessment; (b) the treatment and follow-up dummy indicators into contrasts between the level of the outcome at the beginning of the baseline phase versus the end of the treatment or follow-up phase; and (c) the time-within-phase variables into an indicator of a linear change in a particular phase (see Arntz, Sofi, & van Breukelen [2013] for a similar multilevel models in a multiple baseline design). All effects were considered to be random at Level 2, which allowed capturing between-subjects variations in these effects. In addition, to account for autocorrelation between adjacent sessions, a first-order autoregressive structure was imposed on the covariance matrix for the within-person residuals. To test whether the effects were above and beyond changes in depression levels, all models were also examined while treating BDI scores on each particular assessment as a covariate.
We calculated pseudo $R^2$ for each model according to the recommendations outlined by Peugh (2014; see also Snijders & Bosker, 1999). Specifically, we solved the multilevel equations for each patient on each assessment point for both the unconditioned model (i.e., an intercept only model) and the conditioned model (i.e., a model that includes all predictors) to obtain the predicted outcome on a particular assessment point for these two models. Afterward, we calculated the squared Pearson’s $r$ of the predicted outcome from the unconditioned model and the observed outcome, and the squared Person’s $r$ of the predicted outcome from the conditioned model and the observed outcome. We then used the difference between these two squared Person’s $r$ as an indicator of the variance explained by the conditioned model above and beyond the unconditioned model (i.e., the variance explained by the predictors in the conditioned model).

Finally, the reliable change index (RCI) and clinical significance were computed for the LSAS and SPIN using the method outlined by Jacobson and Truax (1991). Specifically, RCI for each patient was computed as the change in SPIN or LSAS scores from pre- to posttreatment divided by the standard error of the difference of the particular measure (which was calculated based on the scale’s standard deviation and reliability as reported in Heimberg et al. [1999] for the LSAS and Connor et al. [2000] for the SPIN). To meet criteria for clinically significant change, patients had to demonstrate reliable change and also have a posttreatment score lower than 43.49 on the LSAS (determined using Fresco et al.’s reports of clinical and healthy sample means) or lower than 31.1 on the SPIN, following Kocovski, Fleming, Hawley, Huta, and Antony (2013), who used a midpoint between scores of a healthy sample (i.e., a model that includes all predictors) to obtain the predicted outcome from the unconditioned model and the observed outcome.

### Results

Table 3 presents the means and standard deviations of all study variables at five assessment points (i.e., beginning of baseline, end-of-baseline/beginning-of-treatment, end of treatment, 6-month follow-up, and 12-month follow-up). One patient decided to discontinue treatment after 10 sessions. Of the 11 completers, seven (63.64%) did not meet criteria for SAD in the posttreatment interview. The main analyses reported were based on the entire sample (i.e., ITT analyses), but we complemented them with completer analyses (i.e., analyses based on the sample without the patient who dropped out).

### Treatment Adherence

For empathy adherence, the mean score of both raters on the Truax Accurate Empathy scale was 7.32 (pooled $SD = 1.10$) and no ratings fell below 5. Interrater reliability was high and significant, ICC(3,2) = .85, $p > .05$. Regarding adherence to both chair-work tasks, the percentage of therapist talk turns that fell into one of the seven categories of the scales was 85% in the two-chair task and 87% in the empty-chair task. Interrater reliability was high and significant in both tasks, ICC(3,2) = .83, $p < .05$, and ICC(3,2) = .81, $p < .05$, respectively.

### Liebowitz Social Anxiety Scale

A one-way repeated-measures ANOVA, with time as the independent variable and LSAS as the dependent variable, revealed a significant effect for time, $F(4, 44) = 21.42$, $\eta_p^2 = .661$, $p < .001$ (for the completer analysis, $F(4, 40) = 30.08$, $\eta_p^2 = .75$, $p < .001$). For testing specific comparisons between assessment points, we ran post hoc comparisons using Bonferroni correction (i.e., corrected $\alpha = .005$; see Table 3 for LSAS means and SDs). These analyses showed (a) that LSAS scores at the end of the baseline period did not differ significantly from LSAS scores at the beginning of the baseline period, $t(11) = -1.54$, $p = .152$, Cohen’s $d = -1.54$, and at the beginning of treatment assessment, $t(11) = -6.12$, $p < .001$, Cohen’s $d = -2.37$, but were not significantly different from LSAS scores at the 6-month follow-up assessment, $t(11) = 0.04$, $p = .969$, Cohen’s $d = 0.01$ or at the 12-month follow-up assessment, $t(11) = 0.47$, $p = .649$, Cohen’s $d = 0.15$; (c) that LSAS scores at the 6-month follow-up assessment were lower than both LSAS scores at the beginning of baseline period assessment, $t(11) = -4.95$, $p < .001$, Cohen’s $d = -1.38$, and at the beginning of treatment, $t(11) = -5.45$, $p < .001$, Cohen’s $d = -2.03$, but were not significantly different from LSAS scores at the 12-month follow-up assessment, $t(11) = 0.60$, $p = .563$, Cohen’s $d = 0.11$; and (d) that LSAS scores at the 12-month follow-up assessment were lower than both LSAS scores at the beginning of baseline period assessment, $t(11) = -4.93$, $p < .001$, Cohen’s $d = -1.50$, and at the beginning of treatment, $t(11) = -6.61$, $p < .001$, Cohen’s $d = -2.20$.

In sum, we found that patients’ LSAS levels that were assessed after treatment was completed (i.e., the last three assessment points) were significantly lower than LSAS levels that were assessed before the treatment began (i.e., the first two assessment points), with large effect size estimates. In addition, no significant differences were found between the two pretreatment assessments or between the three posttreatment assessments.

### Social Phobia Inventory

As Table 4 demonstrates, multilevel analysis for the SPIN revealed that the treatment and follow-up effects were negative and significant, indicating that at the end of these phases, patients’ SPIN scores were significantly lower than at the beginning of the baseline phase. In addition, time-within-treatment was also negative and significant, indicating that during this phase, there was a linear decline in SPIN scores. In contrast, neither the time-within-baseline nor the time-within-follow-up effects were significant, indicating that there was no linear change in SPIN scores during these phases. The pseudo $R^2$ for the model was .33. When BDI was entered into the model

---

1 The end-of-treatment clinician-administered LSAS score was missing for the client who dropped out of treatment; thus, for this analysis, this client’s LSAS score was carried forward from the beginning of treatment assessment.
Table 3
Means and Standard Deviations for All Outcome Variables at Each Phase for All 12 Patients

<table>
<thead>
<tr>
<th>Phase</th>
<th>LSAS</th>
<th>SPIN</th>
<th>BFNE</th>
<th>FSCRS-IS</th>
<th>FSCRS-HS</th>
<th>FSCRS-RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Beginning of baseline</td>
<td>72.92 (21.40)</td>
<td>43.00 (9.74)</td>
<td>52.92 (4.64)</td>
<td>22.50 (8.66)</td>
<td>3.83 (3.35)</td>
<td>17.83 (7.72)</td>
</tr>
<tr>
<td>2. Beginning of treatment</td>
<td>79.67 (17.21)</td>
<td>41.75 (14.05)</td>
<td>51.67 (7.43)</td>
<td>23.42 (8.07)</td>
<td>4.67 (5.52)</td>
<td>16.92 (7.68)</td>
</tr>
<tr>
<td>3. End of treatment</td>
<td>48.17 (7.57)</td>
<td>28.33 (9.74)</td>
<td>45.25 (10.22)</td>
<td>15.67 (6.39)</td>
<td>2.17 (1.99)</td>
<td>19.92 (7.93)</td>
</tr>
<tr>
<td>4. 6-month follow-up</td>
<td>48.00 (13.81)</td>
<td>24.17 (8.64)</td>
<td>41.58 (10.71)</td>
<td>15.00 (4.43)</td>
<td>1.92 (1.78)</td>
<td>22.00 (6.27)</td>
</tr>
<tr>
<td>5. 12-month follow-up</td>
<td>46.58 (12.52)</td>
<td>25.08 (8.65)</td>
<td>41.50 (10.77)</td>
<td>15.33 (5.82)</td>
<td>2.58 (2.31)</td>
<td>22.17 (7.02)</td>
</tr>
</tbody>
</table>

Note. LSAS = Liebowitz Social Anxiety Scale; SPIN = Social Phobia Inventory; BFNE = Brief Fear of Negative Evaluation; FSCRS-IS = Forms of Self-Criticizing/Self-Reassuring Scale – Inadequate Self Subscale; FSCRS-HS = Forms of Self-Criticizing/Self-Reassuring Scale – Hated-Self Subscale; FSCRS-RS = Forms of Self-Criticizing/Self-Reassuring Scale – Reassuring-Self Subscale.

The end of treatment clinician-administered LSAS score was missing for the client who dropped out of treatment. Therefore, for computing the sample’s mean and standard deviation for this stage, this client’s LSAS score was carried forward from the beginning of treatment assessment.

Brief Fear of Negative Evaluation Scale

As Table 4 demonstrates, the multilevel analysis for the BFNE scale revealed that treatment and follow-up effects were negative and significant, indicating that at the end of these phases, patients’ BFNE scores were significantly lower than at the beginning of the baseline phase. In addition, both the time-within-treatment and time-within-follow-up effects were also negative and significant, indicating that during these phases, there was a linear decline in BFNE scores. In contrast, the time-within-baseline effect was not significant, indicating that there was no linear change in BFNE scores during this phase. The pseudo $R^2$ for the model was .39. When BDI was entered into the model as a covariate, the follow-up ($b = -4.46, SE = 2.27, p = .050, d = -0.93$), time-within-treatment ($b = -8.85, SE = 2.40, p < .001, d = -1.75$), and time-within-baseline effect ($b = -21, SE = 0.11, p = .048$), and time-within-follow-up ($b = -1.88, SE = 0.86, p = .030$) effects remained significant.

Table 4
Results of the Multilevel Modeling Analyses

<table>
<thead>
<tr>
<th>Variable</th>
<th>b(SE)</th>
<th>p</th>
<th>d</th>
<th>b(SE)</th>
<th>p</th>
<th>d</th>
<th>b(SE)</th>
<th>p</th>
<th>d</th>
<th>b(SE)</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>41.82 (2.76)</td>
<td>&lt;.001</td>
<td>52.41 (1.42)</td>
<td>&lt;.001</td>
<td>22.78 (1.74)</td>
<td>&lt;.001</td>
<td>4.03 (.76)</td>
<td>&lt;.001</td>
<td>17.92 (2.09)</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-W-BL</td>
<td>.00 (.35)</td>
<td>989</td>
<td>-.08 (2.99)</td>
<td>.775</td>
<td>.03 (.23)</td>
<td>.882</td>
<td>.06 (.08)</td>
<td>.496</td>
<td>-.36 (1.33)</td>
<td>.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>-13.17 (2.99)</td>
<td>&lt;.001</td>
<td>-1.52</td>
<td>-7.37 (2.50)</td>
<td>.003</td>
<td>-.177</td>
<td>-6.36 (2.09)</td>
<td>.003</td>
<td>-1.93</td>
<td>-1.78 (6.7)</td>
<td>.009</td>
<td>-.90 (.88)</td>
</tr>
<tr>
<td>T-W-T</td>
<td>-.60 (.14)</td>
<td>&lt;.001</td>
<td>-30 (.12)</td>
<td>.009</td>
<td>-30 (.08)</td>
<td>&lt;.001</td>
<td>-.10 (.03)</td>
<td>&lt;.001</td>
<td>.13 (.05)</td>
<td>.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follow-up</td>
<td>-17.40 (3.07)</td>
<td>&lt;.001</td>
<td>-1.93</td>
<td>-11.40 (2.71)</td>
<td>&lt;.001</td>
<td>-2.55</td>
<td>-7.30 (2.18)</td>
<td>&lt;.001</td>
<td>-1.63</td>
<td>-1.45 (7.5)</td>
<td>.054</td>
<td>-.65</td>
</tr>
<tr>
<td>T-W-FU</td>
<td>-1.39 (1.39)</td>
<td>.318</td>
<td>-1.81 (.88)</td>
<td>.042</td>
<td>-.22 (.68)</td>
<td>.744</td>
<td>.17 (.28)</td>
<td>.543</td>
<td>1.13 (.41)</td>
<td>.007</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $p$ values for fixed effects were based on two-tailed $t$ tests with between-within approximation method for computing degrees of freedom; Cohen’s $d$ was calculated as effect size of the change of a phase with respect to baseline; $d$ = the mean outcome difference between baseline and treatment/follow-up phase, derived from the fixed part MLM divided by the standard deviation of the residual outcome variance derived for the random part of the MLM (see Amtz et al., 2013). SPIN = Social Phobia Inventory; BFNE = Brief Fear of Negative Evaluation; FSCRS-IS = Forms of Self-Criticizing/Self-Reassuring Scale – Inadequate Self Subscale; FSCRS-HS = Forms of Self-Criticizing/Self-Reassuring Scale – Hated-Self Subscale; FSCRS-RS = Forms of Self-Criticizing/Self-Reassuring Scale – Reassuring-Self Subscale; SE = standard error; T-W-BL = time within baseline; T-W-T = time within treatment; T-W-F = time within follow-up.
treatment \( (b = -6.97, SE = 2.28, p = .003, d = -2.28) \), follow-up \( (b = -8.26, SE = 2.33, p < .001, d = -1.90) \), and time-within-treatment \( (b = -0.34, SE = 0.09, p < .001) \) effects.

**Forms of Self-Criticizing/Self-Reassuring**

*Scale–Hated-Self*

As Table 4 demonstrates, the multilevel analysis for the Hated-Self scale revealed that the treatment effect was negative and significant, and the follow-up effect was negative and marginally significant, indicating that at the end of this phase patients' Hated-Self scores were lower than at the beginning of the baseline phase. In addition, the time-within-treatment was also negative and significant, indicating that during this phase there was a linear decline in Hated-Self scores. In contrast, neither the time-within-baseline nor the time-within-follow-up effects were significant, indicating that there was no linear change in Hated-Self scores during these phases. The pseudo \( R^2 \) for the model was .08. When BDI was entered into the model as a covariate, the follow-up \( (b = 3.10, SE = 1.12, p = .006, d = 0.46) \), time-within-baseline \( (b = -0.40, SE = 0.12, p = .001) \), and time-within-follow-up \( (b = 1.17, SE = 0.40, p = .004) \) effects remained significant, and the time-within-treatment effect approached significance \( (b = 0.08, SE = 0.04, p = .066) \).

**Forms of Self-Criticizing/Self-Reassuring**

*Scale–Reassuring-Self*

As Table 5 demonstrates, eight of the 11 completers (72.7%) showed a reliable change on the SPIN. Of these eight patients, seven also met criteria for clinical significance, as their posttreatment SPIN scores were lower than 31.1. One patient, although showing a reliable improvement of 16 points on the SPIN, had a posttreatment score of 40, still remaining above the cutoff. Table 5 also shows that all patients showed a reliable improvement on the LSAS. However, only three patients ended treatment with an LSAS score lower than 43.39 and met criteria for clinical significance. Three additional patients had posttreatment LSAS scores between 44 and 47, approaching the clinical cutoff.

**Discussion**

We evaluated the efficacy of EFT for socially anxious patients using a nonconcurrent multiple-baseline design and found that EFT was quite efficacious for this patient group. Of the 11 completers, seven did not meet criteria for SAD at the posttreatment diagnostic interview. Eight patients showed a reliable improvement on the SPIN, and all patients demonstrated a reliable improvement on the LSAS. In addition, seven patients met criteria for a clinically significant change on the SPIN, but only three patients met criteria for clinically significant change on the LSAS. Multilevel modeling analysis with the SPIN and the BFNE, and a one-way ANOVA with the LSAS, using the ITT sample, showed that social anxiety symptoms did not change during the baseline phase, significantly improved during treatment, and then remained improved, or even continued to improve during the 12-month

**Table 5**

| Pretreatment Scores, Posttreatment Scores, Reliable Change, and Clinical Significance for Each of the 11 Completers for the Social Phobia Inventory (SPIN), and for the Liebowitz Social Anxiety Scale (LSAS) |
|---|---|---|---|---|---|---|---|
| Patient | Pretreatment | Posttreatment | Reliable change | Clinical significance | Pretreatment | Posttreatment | Reliable change | Clinical significance |
| 1 | 29 | 24 | No | No | 71 | 50 | Yes | No |
| 2 | 50 | 26 | Yes | Yes | 77 | 60 | No | Yes |
| 3 | 42 | 35 | No | No | 57 | 40 | Yes | Yes |
| 4 | 44 | 30 | Yes | Yes | 95 | 59 | Yes | No |
| 5 | 37 | 12 | Yes | Yes | 88 | 39 | Yes | Yes |
| 6 | 51 | 46 | No | No | 84 | 53 | Yes | No |
| 7 | 42 | 13 | Yes | Yes | 101 | 40 | Yes | Yes |
| 8 | 41 | 31 | Yes | Yes | 71 | 46 | Yes | No |
| 9 | 58 | 30 | Yes | Yes | 98 | 44 | Yes | No |
| 10 | 56 | 40 | Yes | No | 91 | 47 | Yes | No |
| 11 | 56 | 40 | Yes | No | 91 | 47 | Yes | No |
| 12 | 40 | 28 | Yes | Yes | 80 | 57 | Yes | No |
follow-up phase. This pattern of results remained largely the same when depressive symptoms were statistically accounted for.

This is the first study that examined the efficacy of EFT for SAD, and, in essence, is the first study that evaluated EFT for any of the anxiety disorders. EFT has been largely tested with depressed individuals and individuals who had experienced repeated and severe childhood maltreatment, and this study provides initial evidence that anxiety difficulties can also be effectively targeted using EFT. Currently, most empirically supported treatment approaches for anxiety disorders, in general, and for SAD, in particular, are based on exposure and cognitive modification strategies aimed at helping patients down regulate anxiety. The current study provides initial evidence that a treatment that targets different processes, that is, an approach focused on arousing core painful emotions in order to change them with other, more adaptive emotions (assertive anger, grieving, and self-compassion), can also be efficacious. Although studies examining these specific change processes in EFT for SAD are still lacking, this study generally suggests that anxiety symptoms can be reduced via different routes.

The results of this study also show that EFT efficaciously targeted self-criticism and self-reassurance processes that are central in SAD (Cox et al., 2004; Werner et al., 2012). Moderate levels of self-criticism, measured by the Inadequate-Self subscale of the FSCRs, and more severe self-criticism (characterized by self-disgust and wanting to self-injure), measured by the Hated-Self subscale, decreased during the treatment, and gains were maintained during the 12-month follow-up phase. Also, for both measures, the contrasts between the posttreatment point and prebaseline and between post-follow-up and prebaseline were significant. For moderate self-criticism, this pattern remained significant when depressive symptoms were accounted for, but not for severe self-criticism. When depressive symptoms were statistically accounted for, the linear decrease during the therapy phase in Hated-Self scores remained significant, but the contrasts became nonsignificant. Thus, it is possible that the effect of treatment on this particular type of self-criticism is, at least partly, related to depressive symptoms. The more robust effect of the treatment on moderate self-criticism points to the possibility that this type of self-criticism is more directly addressed in EFT.

Self-reassurance increased both during the treatment phase and during the follow-up phase, as indicated by the significant time-within treatment and time-within follow-up slopes. After the treatment phase, the contrast between posttreatment and the beginning of baseline was not significant, but after the follow-up phase, the contrast was significant, indicating that self-reassurance continued to improve after therapy ended. This pattern remained significant while statistically accounting for depressive symptoms. Overall, the findings regarding the effects of EFT on self-criticism and self-reassurance are important, given that these processes have not been systematically examined in previous treatment studies for SAD. Future studies are needed in order to further clarify how self-criticism is resolved and how self-compassion is activated in EFT and in other treatment approaches for SAD.

The findings of this study should be interpreted with the following limitations in mind: First, the design of this study controls for the passage of time, but it does not control for other factors that may have accounted for the benefits of the treatment, such as attention and support from a therapist. From an EFT perspective, such “common factors,” and particularly the relational components specified by Rogers (1951), do play a significant role in the change process, but the assumption is that adding active experiential interventions further deepens emotional processing, which leads to better outcomes. For example, Pos, Greenberg, and Warwar (2009) measured the therapeutic alliance and emotional processing in the beginning, working, and termination phases of EFT for depression, and, using path analysis, found that working phase alliance directly predicted deeper emotional processing and indirectly predicted better outcomes. In another study, Goldman et al. (2006) compared client-centered therapy (e.g., treatment that is based solely on supportive relational Rogerian components) with EFT (same supportive components plus experiential interventions) for depression and found that both treatments were efficacious in improving various outcomes, but that EFT was more efficacious in reducing depressive symptoms. These studies suggest that although relational processes such as empathy that are often termed “common” are essential in EFT, adding more directive interventions designed to activate and transform core maladaptive emotions increases effectiveness. Additional randomized trials with larger samples are needed in order to clarify the role of relational (“common”) factors and specific EFT tasks in the treatment of SAD.

Second, the patients and the therapists were not blind to the study goals and were not blind to the length of the baseline. In that sense, the design of this study is not different from a wait-list design in which patients are “instructed to wait” before therapy begins. Ideally, patients in multiple-baseline designs should not be informed about the beginning point in therapy and should be contacted by their therapist just before the beginning of therapy. Third, the study sample was quite homogenous, comprised of treatment-seeking young adults, and future studies should be conducted with more diverse samples. Fourth, the sampling frequency of the assessment points during the study phases was not even. Specifically, during the baseline period, assessments were conducted every 2 weeks, and during the therapy phase assessments, were conducted every week, before each therapy session. Although multilevel modeling is especially adequate for dealing with data that include different intervals between measurement points (Raudenbush & Bryk, 2002), and the piecewise slopes models enable us to model different trajectories in these different phases of treatment, future studies should use equal assessment intervals across study phases in order to rule out alternative explanations that might emerge from uneven assessments. Fifth, because of budgetary reasons, the MINI was not administered during the 6-month and 12-month follow-up assessments. Future studies would benefit from administering a diagnostic interview in follow-up visits to monitor whether diagnostic status remains stable in response to treatment.

Finally, future studies should assess the theory of change in EFT for SAD. We have already begun this line of work and have found that levels of shame declined, and assertive anger increased, during the treatment, and that assertive anger during chair-work in a given session predicted a reduction in social anxiety symptoms during the following week (Haberman, Shahar, Zilcha-Mano, & Diamond, in press). These findings add support to the central tenets in the EFT change theory regarding SAD.

To summarize, this study provides initial empirical support for EFT as an efficacious treatment for SAD. Future studies are needed in order to expand this line of investigation with larger...
samples, to compare EFT with other empirically supported treatments, and to identify the exact mechanisms leading to change.

References


Received March 6, 2016
Revision received October 18, 2016
Accepted October 21, 2016