Behaviour Research and Therapy 193 (2025) 104825



Contents lists available at ScienceDirect

Behaviour Research and Therapy

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



Behavioral experiments vs. verbal interventions in cognitive therapy for social anxiety disorder: A randomized controlled trial

Cem Yilmaz [a,*], Arwin Nemani [a], Jan Schittenhelm [a], David M. Clark [b], Graham Thew [b], Ulrich Stangier [a]

[a] Goethe University Frankfurt, Department of Psychology, Clinical Psychology and Psychotherapy, Germany
[b] University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Keywords:
Behavioral experiments
Verbal interventions
Cognitive therapy
Cognitive behavioral therapy
Randomized controlled trial
Naturalistic psychotherapy
Dismantling study

ABSTRACT

Objective: The efficacy of cognitive therapy (CT) for social anxiety disorder (SAD) is well documented, although effect sizes (ES) vary significantly across studies. These variations may be attributable to differences in the active components of the treatments. A key component potentially influencing outcomes is the use of behavioral experiments (BEs), as opposed to purely verbal interventions (VIs) for changing social anxiety related negative beliefs. The present trial compared the effectiveness of CT incorporating behavioral experiments (CT-BE) with that of CT focusing more exclusively on verbal interventions (CT-VI). We hypothesized that CT-BE would lead to greater reductions in social anxiety symptoms.

Methods: Sixty patients were randomly assigned to one of two treatment conditions: CT-BE or CT-VI ($n = 30$ per group). The primary outcome was the severity of social anxiety symptoms, which was assessed via the clinician-administered Liebowitz Social Anxiety Scale (LSAS). This measure was also used to compute response. The secondary outcomes included the self-administered Social Phobia Inventory (SPIN), clinician-rated and self-rated depression, and clinician-rated skills in emotion regulation regarding social anxiety. Assessments were conducted at baseline, posttreatment and (for clinician-administered measures) 6-months follow-up.

Results: No significant differences between treatment groups were found for social anxiety outcomes or any other measure at post-treatment, however, a statistical trend in favor of CT-BE was observed for the primary outcome LSAS ($p = .072$). Between-group ES for anxiety outcomes at post-treatment were sizeable ($d = 0.55$ for LSAS, $d = 0.41$ for SPIN), suggesting a potential advantage of CT-BE over CT-VI. Although both treatments showed high overall effectiveness, as reflected in large within-participant ES for anxiety outcomes and high response probabilities (97.7 % for CT-BE, 74.9 % for CT-VI), within-participant ES for anxiety-related outcomes were particularly high in the CT-BE group (up to $d = 2.59$) and substantially exceeded those in the CT-VI group (up to $d = 1.02$). At follow-up, significant group differences emerged for emotion regulation, favoring CT-BE ($p < .001$, $d = 1.19$), indicating a large advantage of BEs in fostering long-term improvements in emotion regulation skills.

Conclusion: The results indicate that behavioral experiments may be an effective component in cognitive therapy for patients with social anxiety disorder, especially facilitating long term emotional regulation skills, in addition to verbal interventions and other experiential exercises. Future research should expand the use of behavioral experiments to larger trials and clarify which mechanisms are involved in the mediating effects of behavioral experiments on outcomes.

Trial registration: German Clinical Trials Register (DRKS), **DRKS00014349**.

1. Introduction

Social anxiety disorder (SAD) is a prevalent mental health condition characterized by intense and persistent fear of social situations, particularly the fear of being negatively evaluated by others (American Psychiatric Association, 2013). SAD is associated with significant distress, impaired social and occupational functioning, an increased risk of comorbidity, and diminished quality of life (e.g. World Mental Health Survey Collaborators et al., 2017).

Individual cognitive behavioral therapy (CBT) is widely regarded as

* Corresponding author. Varrentrappstr. 40-42, 60486, Frankfurt am Main, Germany. Tel.: +49 69 798 23842.
E-mail address: yilmaz@psych.uni-frankfurt.de (C. Yilmaz).

https://doi.org/10.1016/j.brat.2025.104825
Received 2 April 2025; Received in revised form 1 July 2025; Accepted 28 July 2025
Available online 29 July 2025
0005-7967/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the psychological treatment of choice (Carpenter et al., 2018; Kindred et al., 2022; National Institute for Health and Care Excellence [NICE], 2013). However, CBT encompasses a broad range of techniques, and their contribution to the efficacy of treatment can vary substantially. Among the various forms of CBT, cognitive therapy (CT) based on the Clark and Wells (1995) model, has demonstrated particularly high effect sizes in randomized controlled trials (Mayo-Wilson et al., 2014).

While A.T. Beck (1997) emphasized that CT is defined more by its reference to a cognitive model than by specific techniques, traditional techniques of CT typically include verbal interventions (VIs), such as Socratic questioning, as well as behavioral experiments (BEs), both of which are used to modify anxiety related beliefs. Within cognitive approaches to anxiety, BEs involve intentionally engaging with anxiety-provoking situations to test negative beliefs in reality, in contrast with Socratic dialog, which focuses on the rational analysis of cognitive content (D. A. Clark, 2014). The two techniques come together when Socratic dialogue is used prior to BEs to help identify the specific dysfunctional belief that requires correction and to facilitate the design of the BE, and after a BE to maximize the cognitive change resulting from the observed non-occurrence of a predicted negative outcome.

In the treatment of SAD, evidence supporting the effectiveness of BEs is limited. Initial support for this approach came from a case series (Wells et al., 2016), which showed that BEs combined with cognitive interventions led to stronger anxiety reductions and lower beliefs in dysfunctional cognitions than did exposure to BEs based on habituation. In a large randomized controlled trial, D.M. Clark and colleagues (2006) demonstrated that CT was significantly more effective than traditional exposure and applied relaxation. However, CT based on the Clark & Wells model also includes techniques that target cognitive processes that maintain dysfunctional beliefs, such as experiential testing of safety behaviors, attentional training, and video feedback (D. M. Clark, 1999). These techniques have also been shown to be highly effective (McMillan & Lee, 2010) and may contribute substantially to the efficacy of cognitive treatment packages for SAD. Thus, while the cognitive approach based on the Clark and Wells model has been investigated in numerous studies (D. M. Clark et al., 2006, 2023; Ingul et al., 2014; Leichsenring et al., 2013; Mörtberg et al., 2007; Nordahl et al., 2016; Stangier et al., 2003, 2011; Yoshinaga et al., 2016), there is a lack of component analyses clearly demonstrating the effectiveness of BEs.

From a theoretical perspective, the mechanism of BEs is thought to involve disconfirming negative predictions about anticipated aversive outcomes in social situations (D. M. Clark, 2004). Unlike verbal interventions such as Socratic questioning, BEs achieve cognitive change by encouraging individuals to approach social situations, predict the worst that they think might happen and then test their predictions while dropping habitual safety behaviors and/or intentionally breaking excessively strict rules about acceptable behaviors (a procedure often described as "decatastrophizing"). Therefore, a crucial component of BEs—approaching anxiety-provoking situations—is also shared with exposure.

In its traditional form, repeated exposure and prevention of escape/avoidance as crucial components were postulated to enhance habituation and extinction of fear (Foa & Kozak, 1986; Marks, 1978). Although traditional exposure achieved significant symptom reduction in patients with social phobia (Butler et al., 1984; Emmelkamp et al., 1985; Mattick & Peters, 1988) early reports highlighted difficulties in maintaining prolonged exposure situations (Butler, 1985) and cognitive factors interfering with habituation during exposure. In addition, meta-analyses revealed only modest effects sizes and suggested that exposure therapy is less effective than cognitive therapy (Mayo-Wilson et al., 2014; Ougrin, 2011). In contrast to traditional explanations of exposure, modern theories such as Craske's inhibition learning model (Craske et al., 2014) emphasize that expectation violation is a critical component of exposure. In agreement with cognitive theorists, Craske and colleagues also suggest that strategies focusing on information processing, such as removing safety signals, refocusing attention on concrete indicators of rejection, and memory reconsolidation, may enhance the effectiveness of exposure. Consequently, despite essential differences in the theoretical foundations of exposure and BEs, there is growing convergence in the procedural features and explanations of their mechanisms (Hofmann, 2008). Despite substantial conceptual overlap between exposure and behavioral experiments, the two differ in their primary objectives: behavioral experiments aim to disconfirm dysfunctional beliefs by generating experiential evidence, whereas the inhibitory learning model of exposure seeks to develop an inhibitory memory that competes with the original fear memory (Weisman & Rodebaugh, 2018).

However, BEs are not a standard intervention for all therapists taking a broad CBT approach to treating SAD in routine practice settings. This may be due to a lack of knowledge about their efficacy, neglect of this technique in clinical training, and the increased effort required to integrate BEs into the organizational structure of psychotherapy practices. For example, in Germany many clinicians prefer verbal strategies such as automatic thought records or Socratic questioning over BEs (Hoyer et al., 2017). Furthermore, clinicians may be reluctant to implement exposure techniques, due to concerns about potential negative reactions or exacerbating symptoms (Deacon et al., 2013). With regard to BEs in particular, the complexity of social situations may make it difficult for therapists to rule out potential negative outcomes. As a result, there is a need to clarify the effectiveness of BEs and their specific contribution to the efficacy of CT via a dismantling design.

Our study aimed to examine whether CT incorporating BEs (CT-BE) leads to better outcomes in the treatment of SAD than does CT, which relies solely on verbal techniques for cognitive restructuring (CT-VI). To disentangle the effects of BEs from those of other components of CT, both conditions included an initial treatment module aimed at modifying biased information processing. We hypothesized that, compared with CT-VI, CT-BE would result in a higher rate of responders at posttreatment and follow-up. Additionally, we hypothesized that compared with CT-VI, CT-BE would lead to greater improvements in symptom-related outcomes at posttreatment and follow-up. We also expected that the number of BEs would be positively associated with changes in SAD-related outcomes. Furthermore, we examined the effects of the treatments on depressive symptoms and emotion regulation skills with respect to social anxiety.

2. Method

2.1. Design

A mixed model design with two treatment groups (CT-BE and CT-VI) was used. The participants were randomly assigned to one of the two groups. Outcome assessment was conducted by blinded independent raters at baseline, posttreatment and at the 6-month follow-up. Treatment consisted of up to 16 (CT-BE) vs. 25 (CT-VI) sessions; CT-BE sessions tend to be longer to allow experiments to be conducted. The protocol allowed more sessions of CT-VI so the total time of provided therapy was similar (25 × 50 min). The study was designed to detect a moderate effect size ($d = 0.60$). Based on a power analysis with 90 % power on a two-sided test, with $\alpha = .05$ and an expected dropout rate of 10 %, a total sample size of 60 was determined. This study was registered in the German Clinical Trials Register (DRKS) under the identifier DRKS00014349.

2.2. Participant recruitment

The inclusion criteria for participants were as follows: 1) aged 18–70 years, 2) had a primary diagnosis of SAD according to the DSM-5, 3) had no concurrent psychopharmacological or psychotherapeutic treatment, and 4) provided written informed consent. The exclusion criteria were as follows: 1) acute suicidality, 2) active substance abuse or dependence within the past 3 months, 3) severe depression, 4) psychotic disorder, 5)

bipolar disorder, 6) borderline personality disorder, and 7) organic mental disorder or severe medical conditions. Participants were recruited between October 2017 and May 2022 via the outpatient unit at Goethe University Frankfurt, flyers at local sites, SAD self-help-related websites and groups, local radio and television announcements and social media. Depending on how participants learned about the study (via the outpatient unit or by directly contacting the study coordination), potential participants were either first phone-screed by trained MA-level research assistants and subsequently interviewed by administered clinicians or interviewed by latter directly.

Fulfillment of inclusion criteria was ensured prior to treatment through a multi-step diagnostic process. Therapists conducted structured clinical interviews using the German version of the Structured Clinical Interview for DSM-5 – Clinical Version (SCID-5-CV; Beesdo-Baum et al., 2019). In addition, both self-report and clinician-rated diagnostic measures were administered to assess symptom severity and confirm diagnostic impressions (see *Outcome Measures*). Diagnostic results, including comorbidities, were discussed in supervision sessions to ensure diagnostic accuracy and consistency across cases. After successfully meeting the inclusion criteria, participants were randomly allocated to CT-BE or CT-VI at a ratio of 1:1.

All included participants provided written informed consent for the procedures. Fig. 1 displays the CONSORT flowchart.

2.3. Randomization

Two levels of randomization were applied. On the one hand, therapists were randomized to either CT-BE or CT-VI as described above. On the other hand, participants were randomly assigned to a treatment group. Randomization was conducted via a randomization program. Allocation to the treatment of the participants was concealed from participants, raters, and investigators.

2.4. Treatments

CT for SAD comprises three phases: (1) modification of biased information processing, which maintains SAD-related cognitions and beliefs; (2) modification of SAD-related beliefs of negative evaluation by

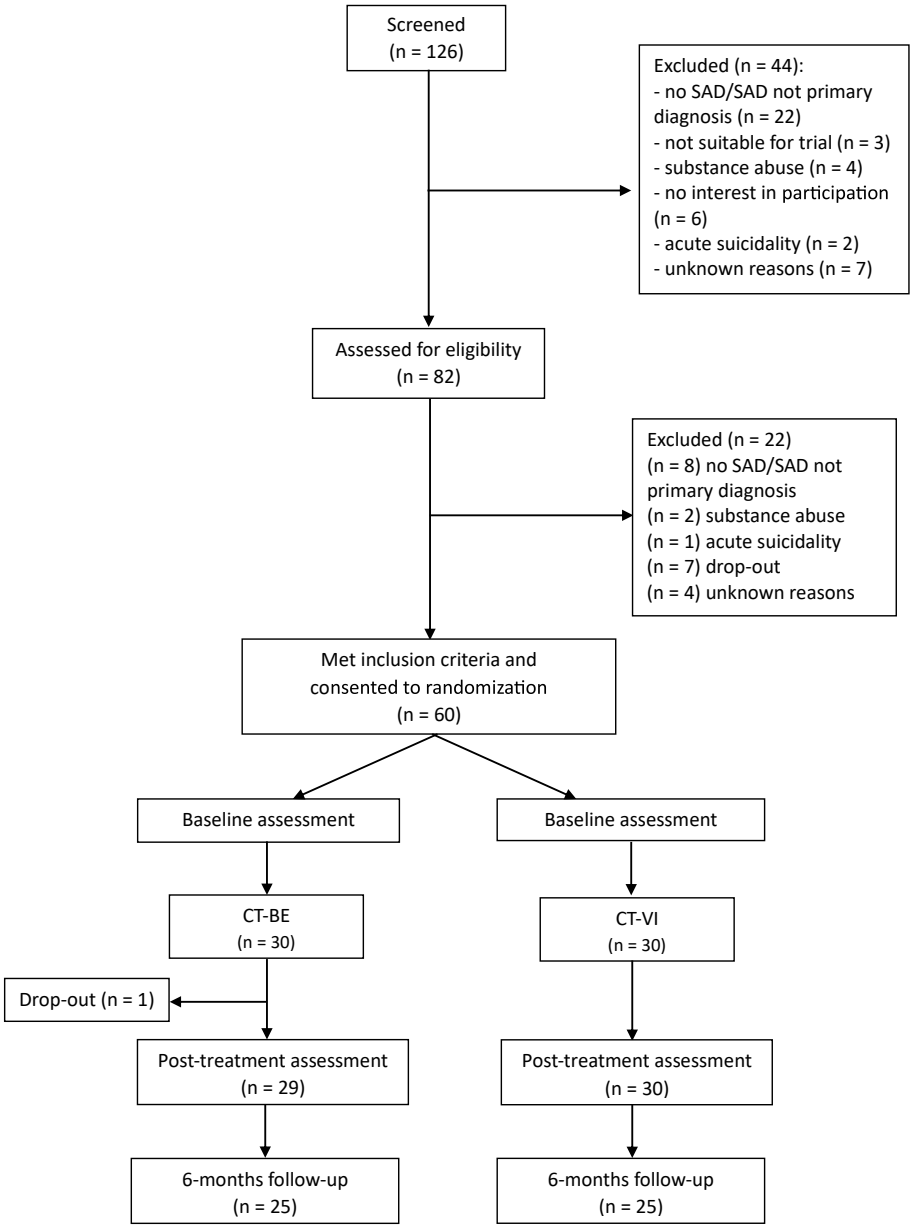


Fig. 1. Flow-chart.

others and negative self-evaluation; and (3) booster sessions and termination of treatment. In this study, two CT-based treatment protocols were compared. The protocols had the first and third phases in common. In phase two, both treatments aimed to change SAD-rated beliefs but differed with respect to the type of interventions (BEs vs. VIs) conducted to do so.

The first phase of treatment comprised four sessions and encompassed the following interventions: deriving an individual cognitive model, conducting an experiential exercise examining the effects of safety behavior and self-focused attention, and providing video-based feedback and attention training. In the second phase, CT-BE focused on BEs (in-session and as homework) but also included verbal interventions such as Socratic questioning of the outcome of BEs. Since BEs require more time, most of the sessions were scheduled for 2 hr, leading to nine CT-BE sessions in this phase. In CT-VI, therapists were instructed to use only verbal interventions within 18 single sessions (guided discovery, automatic thought records, etc.). The third and final stage of treatment consisted of three booster sessions, summarizing the learning points from therapy and encouraging the patient to continue the practice of BEs vs. automatic thought records.

2.5. Therapists and raters

The participants received treatment from trainee clinical psychologists in the latter phase of their training programme. In total, 24 therapists participated in the study (17 females). To ensure a sufficient level of competence, the therapists were trained in the treatment approaches in workshops by the investigator and required to treat one pilot case before conducting the study. Moreover, the therapists received supervision, securing treatment fidelity (90-min group or 60-min single supervision). Supervision was offered after every second double session (CT-BE), or every fourth single session (CT-VI), ensuring a balanced ratio of supervision to applied time of therapy across conditions. Clinical outcomes were assessed by eight independent raters, who were also trainee clinical psychologists who received a half-day workshop from the lead author. However, assessors were not involved in treatment training, supervision or discussion of treatment to ensure blindness.

2.6. Allegiance and adherence

Allegiance was evaluated after therapist training and treatment of a pilot patient via the therapist version of the Reaction to Treatment Questionnaire (Holt & Heimberg, 1990). The instrument consists of four items, with each item being rated on a scale from 1 to 10, with higher scores indicating greater allegiance. For this study, two additional items were included to assess outcome expectations regarding BEs and Vis.

Adherence was ensured by examining video recordings of treatment sessions during supervision. Furthermore, therapists were required to complete session protocols. Adherence scores were computed by calculating the ratio of adherent sessions in phase 2 to the total number of sessions in phase 2. Adherence indices therefore ranged from 0.0 % to 100 %. Adherence to CT-BE in phase 2 was defined as conducting in-session BEs and all other interventions included in the manual if appropriate (Stangier et al., 2016; Warnock-Parkes et al., 2022). At least three in-session BEs in phase 2 were recommended. Imagery rescripting was used in CT-BE if recurring memories of distressing past experiences maintained anxiety related beliefs. Adherence to CT-VI in phase 2 was defined as the application of any of the verbal interventions included in the manual. Nonadherence comprised not only in-session and homework BEs but also interventions focused on cognitive processing, such as attention training, video feedback, modification of safety behavior and imagery rescripting.

2.7. outcome measures

The primary outcome measure for this study was the clinician-administered version of the Liebowitz Social Anxiety Scale (LSAS; (Stangier & Heidenreich, 2005). The LSAS is a widely used instrument in SAD-related research and practice with strong psychometric properties (R. G. Heimberg et al., 1999). The LSAS evaluates fear and avoidance across 24 social situations experienced in the past week via a 4-point scale ranging from 0 (no fear/avoidance) to 3 (severe fear/avoidance). The sum of all the items represents the total LSAS score.

The secondary outcome measures included self-rated social anxiety symptoms, symptoms of depression and skills in emotion regulation in potentially social anxiety-triggering situations.

The German version of the Social Phobia Inventory (SPIN; Sosic et al., 2008) was used to assess patient-rated social anxiety. The SPIN comprises 17 items, gauging fear, avoidance and physiological discomfort experienced in the past week, rated on a 5-point scale from 0 (not at all) to 4 (extremely). The total SPIN score is calculated by summing the responses to the individual items. The inventory has proven to be reliable and valid in assessing social phobia severity (Antony et al., 2006).

Symptoms of depression were on the one hand independently rated via the German version of the Quick Inventory of Depressive Symptomatology (QIDS; Rush et al., 2003). The version used contains 16 items, with each item being rated on a scale from 0 to 3, where higher scores indicate greater symptom severity. The sum of the individual item ratings constitutes the outcome measure. The QIDS demonstrates strong psychometric qualities and is a sensitive measures of symptom severity in depression (Trivedi et al., 2004).

On the other hand, the Beck Depression Inventory (BDI-II (Hautzinger et al., 2005); was applied as a self-report questionnaire for assessing depressive symptom severity. The inventory consists of 21 items, which are rated on a 4-point scale ranging from 0 to 3, with higher scores indicating greater severity of depressive symptoms. The total BDI-II score is the sum of all individual item ratings. The BDI-II is a widely used instrument, with robust psychometric characteristics (Wang & Gorenstein, 2013).

Emotional regulation skills with regard to social anxiety were assessed with the subscale of the Interview for Operationalized Skills Assessment (German Version: OFD; Stenzel et al., 2010). The semi-structured interview assesses the adaptiveness of emotion regulation in five dimensions (identification of emotions, acceptance of emotions, access to strategies of emotion regulation, impulse control and goal-directed behavior as well as appropriate expression of emotions) for four potentially social anxiety-triggering situations. The situations were (1) holding a speech, (2) being the center of attention, (3) experiencing criticism or rejection by a significant other and (4) talking to a stranger. Each dimension was rated on a five-point scale ranging from 1 (“little emotion regulation skills”) to 5 (“high emotion regulation skills”). The individual item ratings for the specific social situations and dimensions were averaged, resulting in a total OFD outcome score. Stenzel et al. demonstrated good psychometric properties for the OFD.

2.8. Response and remission

Response was assessed with the LSAS. On the basis of the findings of von Glischinski et al. (2018), a reduction of at least 29 % in LSAS scores from baseline to posttreatment/follow-up was defined as a response.

For remission, we used the Clinical Global Impression Scale (CGI), modified for individuals with SAD (Stangier et al., 2011; Zaider et al., 2003). The CGI is a regularly used response measure in psychopharmacologic studies and has demonstrated strong psychometric properties (Stangier et al., 2011). The CGI provides information about the current severity of social phobic symptoms, with ratings ranging from 1, meaning “normal” or “not at all ill”, to 7, meaning “among the most severely ill patients”. CGI was assessed at baseline, posttreatment and follow-up. Remission was defined as a CGI score of 1 or 2 at posttreatment/follow-up.

2.9. Statistical analyses

To examine differences in sociodemographic variables between conditions, we performed chi-square tests of independence. When expected cell counts were low, Fisher's exact tests were used instead.

For clinician-rated outcome measures (LSAS, QIDS, and OFD), which were assessed at three timepoints (baseline, post-treatment, and 6-month follow-up), we employed linear mixed-effects models (LMM). Condition (CT-BE vs. CT-VI) and time (post-treatment, follow-up) were included as fixed effects, along with a time-by-condition interaction to estimate treatment effects at each time point. To account for between-subject variability, participant was specified as a random effect. For the self-rated outcome measures (SPIN and BDI), which were assessed at two timepoints (baseline and post-treatment), we conducted standard linear regression models (LRM), predicting post-treatment scores from condition. All models based on continuous data were estimated using restricted maximum likelihood (REML).

For the categorical outcomes of response and remission, we conducted generalized linear mixed-effects models (GLMM). Time (post-treatment and follow-up), condition, and their interaction were included as fixed effects. Participant was specified as a random effect to account for between-subject variability. For response, baseline LSAS and BDI scores were included as covariates. For remission, again, baseline LSAS and BDI scores were included as covariates, along with pre-treatment CGI raw scores as an additional covariate. To improve model stability and prevent convergence issues, all covariates were standardized.

Given that previous research suggests higher baseline severity of social anxiety and depressive symptoms may influence treatment response (e.g. Bauer et al., 2012; Hoyer et al., 2016), we included baseline scores of the LSAS, SPIN, QIDS, and BDI as covariates in all models. For OFD models, baseline OFD scores were included as an additional covariate. For remission analyses, baseline CGI scores were included as an additional covariate. In all GLMMs, covariates were standardized prior to inclusion to improve model stability and reduce convergence issues.

For continuous outcomes, both between-group and within-subject effect sizes were calculated based on Cohen's *d*. Between-group effect sizes were computed by dividing the adjusted mean difference between groups (estimated by the linear mixed effects or regression model) by the unadjusted standard deviation at baseline of the full sample (N). To assess within-subject changes, effect sizes were calculated by dividing the difference between the adjusted mean at post-treatment or follow-up (derived from the linear mixed effects/regression model) and the unadjusted mean at baseline by the unadjusted standard deviation at baseline. This approach aligns with the method described by (Feingold, 2009).

All statistical analyses were conducted using R (version 4.2.2) in R Studio (2024.09.1 + 394) and performed on the intent-to-treat sample. The LMM and GLMM were fitted using the **lme4** package (version 1.1.36; Bates et al., 2003, 2015), and adjusted differences were calculated using the **emmeans** package (version 1.10.7; Lenth, 2017). The LRM were computed with the lm-function from base R.

3. Results

3.1. Participant characteristics at baseline

No significant differences between treatment groups were observed with respect to demographic characteristics or comorbidities (Table 1). The participants were in the age range of 19–61 years. A total of 56.7 % of all participants were female, 60 % were single, 56.7 % had no university degree, and 66.7 % were either employed or studying.

3.2. Dropout and treatment deviation

One participant (CT-BE) dropped out in phase 2 of treatment, resulting in a total dropout rate of 1.7 %. Six participants (two with CT-BE and four with CT-VI) continued therapy after the study protocol or restarted therapy shortly after termination of the study. One CT-VI treatment was finished after eight sessions due to quick remission.

Table 1
Demographic characteristics and comorbidities of participants at baseline.

Characteristic	CT-BE (N = 30)	CT-VI (N = 30)	Statistic	*p*-value
Age in years, *M* (*SD*)	29.8 (10.9)	29.6 (10.1)	$t(58) = 0.10$	0.922
Female, *n* (%)	18 (60)	16 (53)	$\chi^2(1) = 0.27$	0.602
Marital status, *n* (%)				
single	17 (57)	19 (63)	$\chi^2(1) = 0.28$	0.598
relationship or married	10 (33)	6 (20)	$\chi^2(1) = 1.40$	0.237
Education, *n* (%)				
Nonuniversity degree	18 (60)	16 (53)	$\chi^2(1) = 0.27$	0.602
University degree	7 (23)	10 (33)	$\chi^2(1) = 0.74$	0.390
Employment, *n* (%)				
employed	15 (50)	10 (33)	$\chi^2(1) = 1.71$	0.190
student	6 (20)	9 (30)	$\chi^2(1) = 0.80$	0.371
unemployed	2 (7)	3 (10)	*Fisher's exact test*	1.00
Comorbidity, *n* (%)				
No comorbidities	17 (57)	14 (47)	$\chi^2(1) = 1.98$	0.160
Depressive disorder(s)	8 (27)	11 (37)	$\chi^2(1) = 0.22$	0.642
Avoidant Personality disorder	3 (10)	0	*Fisher's exact test*	0.094

Notes. CT-BE = Cognitive Therapy with an emphasis on behavioral experiments; CT-VI = Cognitive Therapy based on verbal interventions.

3.3. Primary and secondary outcomes

Unadjusted mean scores, standard deviations, *t*-test statistics for baseline scores, adjusted differences as well as effect sizes and confidence intervals based on adjusted means for the continuous outcome measures are presented in Table 2.

LMM indicated a trend toward a between-group difference in LSAS scores at post-treatment, favoring CT-BE ($\Delta M = -12.61$, $t(55.1) = -1.83$, $p = .072$), although this did not reach conventional levels of statistical significance. No evidence of a significant difference was found at follow-up ($\Delta M = -7.66$, $t(61.2) = -1.06$, $p = .292$). Between-group effect sizes favored CT-BE over CT-VI at both time points, with a moderate effect at post-treatment ($d = 0.55$) and a small effect at follow-up ($d = 0.33$).

For secondary anxiety measure SPIN, LRM showed no significant difference between CT-BE and CT-VI at post-treatment ($\Delta M = -4.97$, $t(43) = -1.54$, $p = .132$), while the between-group effect size favoring CT-BE was small ($d = 0.41$).

For depression measures QIDS (post-treatment: $\Delta M = -0.51$, $t(57.9) = -0.34$, $p = .736$, $d = 0.10$; follow-up: $\Delta M = -0.48$, $t(61.4) = -0.31$, $p = .757$, $d = 0.09$) and BDI ($\Delta M = -0.11$, $t(43) = -0.04$, $p = .968$, $d = -0.01$), analyses showed no significant differences between treatment groups at any given timepoint.

For OFD, LMM revealed no significant difference between conditions at post-treatment ($\Delta M = 1.61$, $t(59.1) = 1.77$, $p = .083$), but a significant difference at follow-up in favor of CT-BE ($\Delta M = 3.52$, $t(63.8) = 3.66$, $p < .001$). Between-group comparisons on the OFD revealed effect sizes in favor of CT-BE, with a moderate difference at post-treatment ($d = 0.58$) and a large difference at follow-up ($d = 1.19$).

Within-participant analyses regarding baseline to posttreatment changes revealed large improvements in both groups in terms of anxiety-related outcomes, with effect size values of $d = 2.59$ and $d = 2.37$ (LSAS/SPIN) for CT-BE and $d = 1.02$ and $d = 0.97$, respectively, for CT-VI. For OFD, the effect sizes were also high ($d = 1.81$ for CT-BE, $d = 1.74$ for CT-VI), whereas effect sizes for changes in depressive symptoms ranged from small ($d = 0.29$, QIDS, CT-VI), to nearly moderate ($d = 0.48$, BDI,

Table 2
Comparisons between CT-BE and CT-VI on clinical outcomes.

Measure	Time	Unadjusted mean (*SD*) (*n*)		*t*(df), *p*-value	Adjusted difference [95 % CI], *p*-value, *d*
		CT-BE	CT-VI		
LSAS	Pre	85.37 (18.3) (30)	77.68 (26.7) (30)	$t(58) = 1.30$, 0.199	
	Post	43.17 (20.3) (29)	48.03 (26.7) (30)		−12.61 [-26.4, 1.2], 0.072, 0.55
	Follow-Up	42.36 (20.4) (25)	44.84 (29.9) (25)		−7.66 [-22.1, 6.74], 0.292, 0.33
SPIN	Pre	37.61 (9.2) (23)	34.76 (14.3) (29)	$t(50) = 0.83$, 0.411	
	Post	19.04 (12.6) (24)	20.00 (12.0) (29)		−4.97 [-11.5, 1.6], 0.132, 0.41
QIDS	Pre	7.23 (5.0) (30)	6.40 (5.1) (30)	$t(58) = 0.64$, 0.526	
	Post	4.22 (4.3) (27)	4.97 (4.5) (30)		−0.51 [-3.5, 2.5], 0.736, 0.10
	Follow-Up	5.16 (5.2) (25)	5.08 (5.5) (25)		−0.48 [-3.5, 2.6], 0.757, 0.09
BDI	Pre	16.52 (9.4) (23)	12.40 (9.2) (30)	$t(51) = 1.61$, 0.114	
	Post	10.88 (13.0) (24)	7.13 (7.5) (30)		−0.11 [-5.3, 5.1], 0.968, 0.01
OFD	Pre	56.62 (13.4) (29)	56.30 (10.2) (30)	$t(57) = 0.10$, 0.918	
	Post	80.19 (10.0) (26)	75.03 (10.0) (30)		1.61 [-0.21, 3.44], 0.083, 0.58
	Follow-Up	81.98 (12.2) (24)	72.43 (12.6) (23)		3.52 [1.56, 5.45], <0.001, 1.19
CGI	Pre	5.23 (0.7) (30)	5.20 (0.9) (30)	$t(58) = 0.16$, 0.871	
	Post	2.77 (1.1) (30)	3.13 (1.4) (30)		
	Follow-Up	2.64 (1.2) (25)	2.75 (1.3) (24)		

Notes. CT-BE = Cognitive Therapy with an emphasis on behavioral experiments; CT-VI = Cognitive Therapy based on verbal interventions. LSAS = Liebowitz Social Anxiety Scale, SPIN = Social Phobia Inventory, QIDS = Quick Inventory of Depressive Symptomatology, OFD = Interview for Operationalized Skills Assessment, CGI = Clinical Global Impression Scale (CGI), modified for individuals with SAD. Adjusted mean differences for LSAS, QIDS and OFD are based on linear mixed effects models, adjusted for baseline scores. Adjusted mean differences for SPIN and BDI are based on linear regression models, adjusted for baseline scores.

CT-VI), to moderate ($d = 0.56$, QIDS, CT-BE), and up to large ($d = 0.94$, BDI, CT-BE).

Baseline to follow-up within-subject analyses showed also large improvements for LSAS ($d = 2.53$ for CT-BE, $d = 1.16$ for CT-VI) as well as OFD ($d = 2.16$ for CT-BE, $d = 1.50$ for CT-VI). For QIDS, the effect size in the CT-VI group was just below the threshold for a small effect ($d = 0.19$), whereas the effect size in the CT-BE group suggested a trend toward a moderate effect ($d = 0.46$).

3.4. Response and remission

At posttreatment, 24 (CT-BE, $n = 29$, 82.8 %) and 19 (CT-VI, $n = 30$, 63.3 %) participants were identified as responders. At follow-up ($n = 25$ for both conditions), 21 (CT-BE, 84.0 %) and 16 (CT-VI, 64.0 %) participants were determined to be responders.

GLMM indicated that the probability of response was descriptively higher in CT-BE compared to CT-VI at both post-treatment (97.7 %, 95 % CI [61.6 %, 99.9 %] vs. 72.6 %, 95 % CI [33.8 %, 93.2 %]) and follow-up (94.9 %, 95 % CI [49.2 %, 99.7 %] vs. 74.9 %, 95 % CI [34.9 %, 94.3 %]). Correspondingly, the adjusted odds of response were 16.38 times higher in CT-BE than in CT-VI at post-treatment (95 % CI [0.76, 354.00], $p = .075$) and 6.19 times higher at follow-up (95 % CI [0.36, 107.00], $p = .210$), although these differences did not reach statistical significance.

With respect to remission, 12 (CT-BE, $n = 30$, 40.0 %) and 13 (CT-VI, $n = 30$, 43.3 %) participants achieved remission after treatment. At follow-up, 13 (CT-BE, $n = 25$, 52.0 %) and 12 (CT-VI, $n = 24$, 50.0 %) participants experienced remission.

GLMM analyses revealed that the estimated marginal probabilities of remission were descriptively higher in CT-BE group compared to CT-VI at both post-treatment (48.3 %, 95 % CI [2.1 %, 76.6 %] vs. 31.8 %, 95 % CI [12.8 %, 59.7 %]) and follow-up (75.2 %, 95 % CI [38.4 %, 93.6 %] vs. 47.9 %, 95 % CI [22.3 %, 74.7 %]). Accordingly, the adjusted odds of remission were greater in CT-BE than in CT-VI at post-treatment (aOR = 2.01, 95 % CI [0.36, 11.20], $p = .427$) and follow-up (aOR = 3.29, 95 % CI [0.46, 23.80], $p = .238$), although these differences did not reach statistical significance.

3.5. Allegiance and adherence

The level of allegiance was high, with CT-BE therapists reporting an average allegiance score of $M = 8.8$ ($SD = 0.7$) and CT-VI therapists reporting an average allegiance score of $M = 8.1$ ($SD = 0.8$). Although there was no significant difference in allegiance across treatments ($t(22) = 2.05$, $p = .052$), there was a strong trend toward higher allegiance in the CT-BE group.

Adherence as assessed by the intervention checklist revealed a high level of adherence. For CT-BE, the proportion of treatment sessions that were assessed as adherent ranged from 77.9 % to 100 % per therapist, with an average of 98.9 % ($SD = 4.7$). The number of CT-BE courses of treatment with the recommended amount of in-session BEs (at least 3) was 26 (89.7 %), with an average of $M = 5.9$ ($SD = 2.3$). For CT-VI, adherence scores ranged from 43.0 % to 100 %, with an average of 85.2 % ($SD = 17.7$). The number of CT-VI patients whose therapist conducted at least one in-session BE was 13 (43.3 %), with an average of $M = 3.2$ ($SD = 2.5$).

Interestingly, adherence to CT-BE was not significantly correlated with pre/post-changes in LSAS scores ($r = 0.33$, $p = .083$), whereas adherence with CT-VI was negatively correlated with pre/post-changes in LSAS scores ($r = -0.37$, $p = .044$).

The number of conducted in-session BEs showed a strong statistical trend towards correlation with pre/post-changes in LSAS scores for CT-BE ($r = 0.37$, $p = .052$). For the whole sample this correlative relationship was significant (r = 0.36, $p = .005$). The correlation between the nonadherent use of BEs (in-session BEs in phase 2 in CT-VI) and LSAS outcomes, however, was not significant ($r = 0.14$, $p = .459$).

For SPIN, pre/post-changes significantly correlated with adherence to CT-BE ($r = 0.56$, $p = .007$), but not with adherence to CT-VI ($r = -0.25$, $p = .189$). Also, pre/post-changes in SPIN scores correlated significantly with the number of in-session BEs for CT-BE ($r = 0.55$, $p = .008$). The nonadherent use of BEs and changes in SPIN scores from baseline to post-treatment was not significant ($r = 0.25$, $p = .190$).

3.6. Post-hoc analysis

In view of the fact that therapists conducted BEs with 43.3 % of patients in the CT-VI condition, we recalculated between-group effect sizes for the primary outcome (LSAS) after excluding these individuals from the CT-VI group. Compared to the original intention-to-treat analysis, effect sizes increased at both post-treatment ($d = 0.55$ vs. $d = 0.64$) and follow-up, where the effect exceeded the threshold for a moderate effect ($d = 0.33$ vs $d = 0.53$). The small sample size for CT-VI group that used no BEs ($n = 17$) meant that formal significance testing was inappropriate.

4. Discussion

The present trial aimed to determine the effectiveness of BEs in the treatment of SAD. Thus, a dismantling design was used to compare cognitive therapy focusing on behavioral experiments (CT-BE) with cognitive therapy focusing on verbal interventions (CT-VI) after an initial common phase of experiential exercises targeting cognitive processes, such as the self-focused attention and safety behaviors experiment, video-feedback and training in externally focused attention for use in social situations. Verbal interventions were chosen as a control condition since this type of intervention is very common in clinical practice.

Neither LMM on clinician-rated severity of social anxiety (primary outcome LSAS, $p = .072$), nor LRM on patient-rated social anxiety (SPIN, $p = .132$) showed a significant difference between treatment conditions at post-treatment. Nonetheless, with the LSAS analysis indicating a statistical trend toward significance and between-group effect sizes favoring CT-BE ($d = 0.55$ for LSAS, $d = 0.41$ for SPIN), the results suggest a potential additional benefit of incorporating BEs for reducing social anxiety symptoms. Within-subject effect sizes were large for both measures in both conditions, yielding high effectiveness of both treatments, but were substantially higher in CT-BE ($d = 2.59$ for LSAS, $d = 2.37$ for SPIN) than in CT-VI ($d = 1.02$ for LSAS, $d = 0.97$ for SPIN). Within-participant analyses from baseline to follow-up of LSAS showed that treatment effects were maintained ($d = 2.53$ for CT-BE, $d = 1.16$ for CT-VI). Interestingly, the effect sizes reached by CT-BE are comparable to the studies conducted by Clark and colleagues (2003, 2006) and significantly greater than those achieved in several recent studies on CT for SAD in Germany, which put less emphasis on BEs (Leichsenring et al., 2013; Stangier et al., 2003, 2011). Thus, although statistical significance was not achieved, findings based on effect sizes indicate that BEs may be an effective component in CT and improve treatment for SAD. This was supported by the post-hoc analysis that showed an increase in BE related effect size when patients whose therapists had conducted in session BEs were excluded from the CT-VI group.

On the basis of the LSAS, rates, probabilities and odds ratios for response were computed for post-treatment and follow-up. Response rates were high for both treatments (82.8 % for CT-BE and 63.3 % for CT-VI at posttreatment, 84.0 % and 64.0 %, respectively, at follow-up). At post-treatment, the probability of achieving response was descriptively higher in CT-BE (97.7 %) compared to CT-VI (72.6 %) with an adjusted odds ratio of 16.38 ($p = .075$). A similar pattern was observed at follow-up, where response probabilities remained high in CT-BE (94.9 %) compared to CT-VI (74.9 %) with an adjusted odds ratio of 6.19 ($p = .210$). Although these differences did not reach statistical significance, the consistently higher response probabilities and odds ratios favoring CT-BE suggest that incorporating BEs may enhance the likelihood of meaningful symptom improvement.

The remission rates at posttreatment (40.0 % for CT-BE, 43.3 % for CT-VI) and follow-up (52.0 % for CT-BE, 50.0 % for CT-VI) are in line with rates reported in previous studies (Loerinc et al., 2015; Springer et al., 2018). Controlling for baseline scores of social anxiety, clinical global impression and depressive symptoms, GLMMs revealed probabilities of remission that descriptively differ from the observed remission rates in favor of CT-BE (48.3 % for CT-BE and 31.8 % for CT-VI at posttreatment, 75.2 % and 47.9 %, respectively, at follow-up).

The assessment of treatment integrity was based on clinical training workshops, different chapters from a treatment manual (Stangier et al., 2016) and checklists of interventions completed by the therapists after each treatment session. Our results indicated that, in general, the content of the majority of sessions was in line with the respective treatment manuals. However, found that therapists used some BEs when treating 43.3 % of patients in the CT-VI condition. Since the use of BEs also correlated with a positive treatment outcome, it is possible our estimate of between treatment effects was reduced by a blurring of the procedural differences between treatment conditions. This assumption is also supported by the post-hoc analysis, which revealed higher between-group effect sizes in favor of CT-BE for post-treatment and follow-up, when CT-VI courses of treatment, in which BEs were conducted, were omitted. A possible explanation for this lack of adherence may be given by the data on allegiance. Despite the randomization of therapists to both treatment conditions, there was a strong tendency toward lower allegiance in the CT-VI condition. Thus, contrary to the instructions, therapists might have used BEs to deliver what they considered to be the most appropriate treatment. BE is a common adjunct to cognitive restructuring, particularly in terms of negative self-evaluations and low self-esteem (Fennell, 2007).

Although the identification of effective components is a high priority in research (Kaczkurkin & Foa, 2015), only a few attempts have been made to dismantle studies of SAD. In an early study, Hope et al. (1995) compared group CBT with exposure without cognitive restructuring and reported only mixed results with respect to different outcome measures. Our goal to dismantle the effect of BEs was impaired by the shared first treatment phase, which comprised the collaborative development of a cognitive model and experiential interventions to modify SAD-related cognitive processes. There is clear evidence that the self-focused attention and safety behaviors manipulation (McManus et al., 2009; Schreiber et al., 2015), attentional training (Knowles et al., 2016; McEvoy & Perini, 2009; Wells et al., 1997) and the video feedback (Rodebaugh et al., 2010) used in phase one are effective treatment components. It is likely they contributed to the large within effect sizes observed with both treatments. Unfortunately, we did not include an intermediate assessment after phase one to disentangle the effects of the first and second treatment phases.

In addition to social anxiety, we explored the effects of CT-BE and CT-VI on depressive symptoms. While there were no significant differences between CT-BE and CT-VI for changes in either clinician-rated (QIDS), or self-rated (BDI) depression outcomes, descriptive scores and within-subject effect sizes indicate that both treatments were effective in reducing depressive symptoms. This is in line with current research, demonstrating that transdiagnostic interventions such BEs and Vis are effective for treating anxiety with comorbid depression (Bentley et al., 2021). Simultaneously, within-subject effect sizes are descriptively in favor for CT-BE, with almost moderate ($d = 0.46$ for QIDS, baseline to follow-up), moderate ($d = 0.56$ for QIDS, baseline to post-treatment) and large ($d = 0.94$, BDI) effect sizes, compared to negligible ($d = 0.19$ for QIDS, baseline to follow-up), small ($d = 0.29$ for QIDS, baseline to post-treatment) and nearly moderate ($d = 0.48$, BDI) effect size values for CT-VI. These outcomes suggest that CT-BE may have provided a slightly stronger reduction in depressive symptoms over time.

The results for emotion regulation skills, as measured by the OFD, yielded also high within-participant effect size values for both treatment groups for both post-treatment ($d = 1.81$ for CT-BE, $d = 1.74$ for CT-VI), and follow-up ($d = 2.16$ for CT-BE, $d = 1.50$ for CT-VI), which suggests that both treatments may be effective in terms of promoting emotion regulation. This supports the finding of a recent study on internet-delivered CBT for SAD patients, which revealed that emotion regulation may be an important mediator of social anxiety reduction (Garke et al., 2024). Regarding treatment comparisons, results showed a trend toward a group difference at post-treatment ($p = .083$, $d = 0.58$), with a clear and substantial advantage for CT-BE observed at follow-up ($p < .001$, $d = 1.19$). This strong effect size suggests that BEs promote more durable improvements in emotion regulation, extending beyond the treatment period. One possible explanation is that BEs and VIs may target different aspects of emotion regulation: whereas VIs primarily focus on cognitive re-evaluation of social situations, BEs encourage direct engagement in functional social behaviors, such as active social interaction instead of relying on safety behaviors (Ginat-Frolich et al., 2024; Shuman, 2013). This behavioral engagement may foster more lasting improvements in the ability to identify, accept, and effectively manage aversive emotions in social contexts.

Our results support the integration of BEs into standard care for SAD. One important implication is the need for more explicit descriptions of BE procedures in research reports. It is also essential to provide greater detail in the description of exposure treatments, given the considerable overlap between exposure based on inhibitory learning and BEs (Craske et al., 2014). Furthermore, to address previously mentioned barriers in clinical practice, the principles underlying the design of BEs should be incorporated into clinical training. This includes opportunities for self-practice and self-reflection (Bennett-Levy, 2003) as well as the dissemination of comprehensive treatment manuals (Rouf, 2004). Developing the therapeutic skills necessary to competently deliver BEs is crucial for the effective treatment of SAD (Ginzburg et al., 2012).

In clinical practice, it is important to recognize that the implementation of BEs must be embedded within a cognitive case conceptualization (D. M. Clark & Wells, 1995). First, because BEs are designed to test the beliefs that maintain social anxiety, identifying these individual beliefs is crucial to their effectiveness. Additionally, since such beliefs are translated into specific expectancies within particular situations, it is essential to define clear and precise criteria for belief confirmation (Warnock-Parkes et al., 2022). Second, it is critically important to minimize factors that interfere with belief disconfirmation during BEs, including the use of safety behaviors and self-focused attention (D. M. Clark & McManus, 2002; Schreiber et al., 2015). Consequently, BEs should be grounded in information obtained beforehand through a personalized, collaboratively developed cognitive model (D. M. Clark & Wells, 1995). Third, because social interactions are often ambiguous, the outcome of a BE may not always be conclusive. Moreover, post-event processing can activate negative memories and amplify negative interpretations, potentially obscuring disconfirming evidence (Flynn & Yoon, 2025). Therefore, guiding the patient through a structured reflection on the outcome of the BE is a critical step in facilitating functional processing of the experience. This process ressembles much more the collaborative process of guided discovery than the didactic strategy of Socratic questioning based on logic (Kazantzis et al., 2018).

Several limitations of this study should be acknowledged. First, the therapists were trainees who had not treated many patients with SAD. Second, the sample size was relatively small, which may limit the statistical power for secondary outcome measures. Third, adherence was only measured by check-ups completed by the therapists; therefore, some uncertainty remains as to whether treatment protocols were followed properly. Fourth, as opposed to adherence, therapist competence was not measured, which also leads to uncertainty regarding the proper conduct of interventions, especially BEs. Moreover, there was substantial overlap in interventions between treatment conditions, and although not encouraged, CT-VI therapists also made use of BEs, which limits internal validity. Furthermore, as mentioned above, we did not include an intermediate measurement point after the initial phase of the treatment, which was identical for both conditions. As a result, the relative contribution of the first treatment phase to the overall effect remains unclear, complicating the disentanglement of the effects of the second phase.

Despite these limitations, several methodological strengths of the study should be noted, including the application of an effective comparison group, the randomization of therapists to treatment conditions, the use of independent and blind assessments, the use of both clinician and patient ratings, and high external validity with respect to practice settings.

In conclusion, this trial compared the effectiveness of cognitive therapy, including behavioral experiments, with that of cognitive therapy, focusing on verbal interventions for social anxiety disorder. Both treatments were highly effective, but the use of behavioral experiments may have an additional beneficial effect on the severity of social anxiety and increase the likelihood of response. Furthermore, behavioral experiments prove to have an effect on long term emotion regulation skills. Future research should clarify the active ingredients of verbal interventions and perform behavioral experiments, particularly with larger samples. However, our results indicate that behavioral experiments may be an effective component of cognitive Therapy for patients with social anxiety disorder.

CRediT authorship contribution statement

Cem Yilmaz: Writing – original draft, Project administration, Investigation, Formal analysis, Data curation. **Arwin Nemani:** Writing – review & editing, Formal analysis. **Jan Schittenhelm:** Writing – review & editing, Investigation, Formal analysis. **David M. Clark:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Graham Thew:** Writing – review & editing, Validation, Formal analysis. **Ulrich Stangier:** Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT.com and deepl.com in order to improve readability and language. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known conflict of interest with any commercial or other associations in connection with the submitted article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2025.104825.

Data availability

Data will be made available on request.

References

American Psychiatric Association. (2013). *Diagnostic and statistial manual of mental disorders* (5th ed.). Author.

Antony, M. M., Coons, M. J., McCabe, R. E., Ashbaugh, A., & Swinson, R. P. (2006). Psychometric properties of the social phobia inventory: Further evaluation. *Behaviour Research and Therapy, 44*(8), 1177–1185. https://doi.org/10.1016/j.brat.2005.08.013

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2003). lme4: Linear mixed-effects models using "eigen" and S4 (p. 1.1-37) https://doi.org/10.32614/CRAN.package.lme4.

Bauer, I., Wilansky-Traynor, P., & Rector, N. A. (2012). Cognitive-behavioral therapy for anxiety disorders with comorbid depression: A review. *International Journal of Cognitive Therapy, 5*(2), 118–156. https://doi.org/10.1521/ijct.2012.5.2.118

Beck, A. T. (1997). The past and future of cognitive therapy. *The Journal of Psychotherapy Practice and Research, 6*(4), 276–284.

Beesdo-Baum, K., Zaudig, M., & Wittchen, H. U. (2019). *SCID-5-CV: strukturiertes klinisches Interview für DSM-5-Störungen—Klinische Version: Deutsche Bearbeitung des Structured Clinical Interview for DSM-5 Disorders—Clinician version von Michael B. First, Janet B.W. Williams, Rhonda S. Karg, Robert L. Spitzer, 1*. Hogrefe.

Bennett-Levy, J. (2003). Mechanisms of change in cognitive therapy: The case of automatic thought records and behavioural experiments. *Behavioural and Cognitive Psychotherapy, 31*(3), 261–277. https://doi.org/10.1017/S1352465803003035

Bentley, K. H., Bernstein, E. E., Wallace, B., & Mischoulon, D. (2021). Treatment for anxiety and comorbid depressive disorders: Transdiagnostic cognitive-behavioral strategies. *Psychiatric Annals, 51*(5), 226–230. https://doi.org/10.3928/00485713-20210414-01

Butler, G. (1985). Exposure as a treatment for social phobia: Some instructive difficulties. *Behaviour Research and Therapy, 23*(6), 651–657. https://doi.org/10.1016/0005-7967(85)90060-9

Butler, G., Cullington, A., Munby, M., Amies, P., & Gelder, M. (1984). Exposure and anxiety management in the treatment of social phobia. *Journal of Consulting and Clinical Psychology, 52*(4), 642–650. https://doi.org/10.1037/0022-006X.52.4.642

Carpenter, J. K., Andrews, L. A., Witcraft, S. M., Powers, M. B., Smits, J. A. J., & Hofmann, S. G. (2018). Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials. *Depression and Anxiety, 35*(6), 502–514. https://doi.org/10.1002/da.22728

Clark, D. M. (1999). Anxiety disorders: Why they persist and how to treat them. *Behaviour Research and Therapy, 37*, S5–S27. https://doi.org/10.1016/S0005-7967(99)00048-0

Clark, D. M. (2004). Developing new treatments: On the interplay between theories, experimental science and clinical innovation. *Behaviour Research and Therapy, 42*(9), 1089–1104. https://doi.org/10.1016/j.brat.2004.05.002

Clark, D. A. (2014). Cognitive restructuring. In *, 1–3. The wiley handbook of cognitive behavioral therapy* (pp. 23–44). Wiley Blackwell.

Clark, D. M., Ehlers, A., Hackmann, A., McManus, F., Fennell, M., Grey, N., Waddington, L., & Wild, J. (2006). Cognitive therapy versus exposure and applied relaxation in social phobia: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 74*(3), 568–578. https://doi.org/10.1037/0022-006X.74.3.568

Clark, D. M., Ehlers, A., McManus, F., Hackmann, A., Fennell, M., Campbell, H., Flower, T., Davenport, C., & Louis, B. (2003). Cognitive therapy versus fluoxetine in generalized social phobia: A randomized placebo-controlled trial. *Journal of Consulting and Clinical Psychology, 71*(6), 1058–1067. https://doi.org/10.1037/0022-006X.71.6.1058

Clark, D. M., & McManus, F. (2002). Information processing in social phobia. *Biological Psychiatry, 51*(1), 92–100. https://doi.org/10.1016/S0006-3223(01)01296-3

Clark, D. M., & Wells, A. A. (1995). A cognitive model of social phobia. In R. Heimberg, M. R. Liebowitz, D. A. Hope, & F. R. Schneier (Eds.), *Social phobia: Diagnosis, assessment and treatment* (pp. 69–93). Guilford Press.

Clark, D. M., Wild, J., Warnock-Parkes, E., Stott, R., Grey, N., Thew, G., & Ehlers, A. (2023). More than doubling the clinical benefit of each hour of therapist time: A randomised controlled trial of internet cognitive therapy for social anxiety disorder. *Psychological Medicine, 53*(11), 5022–5032. https://doi.org/10.1017/S0033291722002008

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy, 58*, 10–23. https://doi.org/10.1016/j.brat.2014.04.006

Deacon, B. J., Lickel, J. J., Farrell, N. R., Kemp, J. J., & Hipol, L. J. (2013). Therapist perceptions and delivery of interoceptive exposure for panic disorder. *Journal of Anxiety Disorders, 27*(2), 259–264. https://doi.org/10.1016/j.janxdis.2013.02.004

Emmelkamp, P. M. G., Mersch, P.-P., Vissia, E., & Van Der Helm, M. (1985). Social phobia: A comparative evaluation of cognitive and behavioral interventions. *Behaviour Research and Therapy, 23*(3), 365–369. https://doi.org/10.1016/0005-7967(85)90015-4

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods, 14*(1), 43–53. https://doi.org/10.1037/a0014699

Fennell, M. J. V. (2007). Low self-esteem. In N. Kazantzis, & L. LʼAbate (Eds.), *Handbook of homework assignments in psychotherapy* (pp. 297–314). Springer US. https://doi.org/10.1007/978-0-387-29681-4_18.

Flynn, A. J., & Yoon, K. L. (2025). Post-event processing in social anxiety: A scoping review. *Journal of Anxiety Disorders, 109*, Article 102947. https://doi.org/10.1016/j.janxdis.2024.102947

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin, 99*(1), 20–35.

Garke, M.Å., Hentati Isacsson, N., Kolbeinsson, Ö., Hesser, H., & Månsson, K. N. T. (2024). Improvements in emotion regulation during cognitive behavior therapy predict subsequent social anxiety reductions. *Cognitive Behaviour Therapy*, 1–18. https://doi.org/10.1080/16506073.2024.2373784

Ginat-Frolich, R., Gilboa-Schechtman, E., Huppert, J. D., Aderka, I. M., Alden, L. E., Bar-Haim, Y., Becker, E. S., Bernstein, A., Geva, R., Heimberg, R. G., Hofmann, S. G., Kashdan, T. B., Koster, E. H. W., Lipsitz, J., Maner, J. K., Moscovitch, D. A., Philippot, P., Rapee, R. M., Roelofs, K., … Wieser, M. J. (2024). Vulnerabilities in social anxiety: Integrating intra- and interpersonal perspectives. *Clinical Psychology Review, 109*, Article 102415. https://doi.org/10.1016/j.cpr.2024.102415

Ginzburg, D. M., Bohn, C., Höfling, V., Weck, F., Clark, D. M., & Stangier, U. (2012). Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy, 50*(12), 747–752. https://doi.org/10.1016/j.brat.2012.09.001

Hautzinger, M., Keller, F., & Kühner, C. (2005). Beck depressions-inventar (BDI-II). *Harcourt Test Services*.

Heimberg, R. G., Horner, K. J., Juster, H. R., Safren, S. A., Brown, E. J., Schneier, F. R., & Liebowitz, M. R. (1999). Psychometric properties of the liebowitz social anxiety scale. *Psychological Medicine, 29*(1), 199–212. https://doi.org/10.1017/S0033291798007879

Hofmann, S. (2008). Cognitive processes during fear acquisition and extinction in animals and humans: Implications for exposure therapy of anxiety disorders. *Clinical Psychology Review, 28*(2), 199–210. https://doi.org/10.1016/j.cpr.2007.04.009

Holt, C. S., & Heimberg, R. G. (1990). The reaction to treatment questionnaire: Measuring treatment credibility and outcome expectancies. *The behavior therapist*.

Hope, D. A., Heimberg, R. G., & Bruch, M. A. (1995). Dismantling cognitive-behavioral group therapy for social phobia. *Behaviour Research and Therapy, 33*(6), 637–650. https://doi.org/10.1016/0005-7967(95)00013-N

Hoyer, J., Čolić, J., Pittig, A., Crawcour, S., Moeser, M., Ginzburg, D., Lin, J., Wiltink, J., Leibing, E., & Stangier, U. (2017). Manualized cognitive therapy versus cognitive-behavioral treatment-as-usual for social anxiety disorder in routine practice: A cluster-randomized controlled trial. *Behaviour Research and Therapy, 95*, 87–98. https://doi.org/10.1016/j.brat.2017.05.012

Hoyer, J., Wiltink, J., Hiller, W., Miller, R., Salzer, S., Sarnowsky, S., Stangier, U., Strauss, B., Willutzki, U., & Leibing, E. (2016). Baseline patient characteristics predicting outcome and attrition in cognitive therapy for social phobia: Results from a large multicentre trial. *Clinical Psychology & Psychotherapy, 23*(1), 35–46. https://doi.org/10.1002/cpp.1936

Ingul, J. M., Aune, T., & Nordahl, H. M. (2014). A randomized controlled trial of individual cognitive therapy, group cognitive behaviour therapy and attentional placebo for adolescent social phobia. *Psychotherapy and Psychosomatics, 83*(1), 54–61. https://doi.org/10.1159/000354672

Kaczkurkin, A. N., & Foa, E. B. (2015). Cognitive-behavioral therapy for anxiety disorders: An update on the empirical evidence. *Dialogues in Clinical Neuroscience, 17*(3), 337–346. https://doi.org/10.31887/DCNS.2015.17.3/akaczkurkin

Kazantzis, N., Beck, J. S., Clark, D. A., Dobson, K. S., Hofmann, S. G., Leahy, R. L., & Wing Wong, C. (2018). Socratic dialogue and guided discovery in cognitive behavioral therapy: A modified Delphi panel. *International Journal of Cognitive Therapy, 11*(2), 140–157. https://doi.org/10.1007/s41811-018-0012-2

Kindred, R., Bates, G. W., & McBride, N. L. (2022). Long-term outcomes of cognitive behavioural therapy for social anxiety disorder: A meta-analysis of randomised controlled trials. *Journal of Anxiety Disorders, 92*, Article 102640. https://doi.org/10.1016/j.janxdis.2022.102640

Knowles, M. M., Foden, P., El-Deredy, W., & Wells, A. (2016). A systematic review of efficacy of the attention training technique in clinical and nonclinical samples. *Journal of Clinical Psychology, 72*(10), 999–1025. https://doi.org/10.1002/jclp.22312

Leichsenring, F., Salzer, S., Beutel, M. E., Herpertz, S., Hiller, W., Hoyer, J., Huesing, J., Joraschky, P., Nolting, B., Poehlmann, K., Ritter, V., Stangier, U., Strauss, B., Stuhldreher, N., Tefikow, S., Teismann, T., Willutzki, U., Wiltink, J., & Leibing, E. (2013). Psychodynamic therapy and cognitive-behavioral therapy in social anxiety disorder: A multicenter randomized controlled trial. *American Journal of Psychiatry, 170*(7), 759–767. https://doi.org/10.1176/appi.ajp.2013.12081125

Lenth, R. V. (2017). Emmeans: Estimated marginal means, Aka least-squares means (p. 1.11.1) https://doi.org/10.32614/CRAN.package.emmeans.

Loerinc, A. G., Meuret, A. E., Twohig, M. P., Rosenfield, D., Bluett, E. J., & Craske, M. G. (2015). Response rates for CBT for anxiety disorders: Need for standardized criteria. *Clinical Psychology Review, 42*, 72–82. https://doi.org/10.1016/j.cpr.2015.08.004

Marks, I. (1978). Behavioral psychotherapy of adult neurosis. In S. L. Garfield, & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change: An empirical analysis* (2nd ed., pp. 493–547). Wiley.

Mattick, R. P., & Peters, L. (1988). Treatment of severe social phobia: Effects of guided exposure with and without cognitive restructuring. *Journal of Consulting and Clinical Psychology, 56*(2), 251–260. https://doi.org/10.1037/0022-006X.56.2.251

Mayo-Wilson, E., Dias, S., Mavranezouli, I., Kew, K., Clark, D. M., Ades, A. E., & Pilling, S. (2014). Psychological and pharmacological interventions for social anxiety disorder in adults: A systematic review and network meta-analysis. *The Lancet Psychiatry, 1*(5), 368–376. https://doi.org/10.1016/S2215-0366(14)70329-3

McEvoy, P. M., & Perini, S. J. (2009). Cognitive behavioral group therapy for social phobia with or without attention training: A controlled trial. *Journal of Anxiety Disorders, 23*(4), 519–528. https://doi.org/10.1016/j.janxdis.2008.10.008

McManus, F., Clark, D. M., Grey, N., Wild, J., Hirsch, C., Fennell, M., Hackmann, A., Waddington, L., Liness, S., & Manley, J. (2009). A demonstration of the efficacy of two of the components of cognitive therapy for social phobia. *Journal of Anxiety Disorders, 23*(4), 496–503. https://doi.org/10.1016/j.janxdis.2008.10.010

McMillan, D., & Lee, R. (2010). A systematic review of behavioral experiments vs. exposure alone in the treatment of anxiety disorders: A case of exposure while wearing the emperor's new clothes? *Clinical Psychology Review, 30*(5), 467–478. https://doi.org/10.1016/j.cpr.2010.01.003

Mörtberg, E., Clark, D. M., Sundin, Ö., & Åberg Wistedt, A. (2007). Intensive group cognitive treatment and individual cognitive therapy vs. treatment as usual in social phobia: A randomized controlled trial. *Acta Psychiatrica Scandinavica, 115*(2), 142–154. https://doi.org/10.1111/j.1600-0447.2006.00839.x

National Institute for Health and Care Excellence [NICE]. (2013). *Social anxiety disorder: Recognition, assessment and treatment* (CG159). https://www.nice.org.uk/guidance/cg159.

Nordahl, H. M., Vogel, P. A., Morken, G., Stiles, T. C., Sandvik, P., & Wells, A. (2016). Paroxetine, cognitive therapy or their combination in the treatment of social anxiety disorder with and without avoidant personality disorder: A randomized clinical trial. *Psychotherapy and Psychosomatics, 85*(6), 346–356. https://doi.org/10.1159/000447013

Ougrin, D. (2011). Efficacy of exposure versus cognitive therapy in anxiety disorders: Systematic review and meta-analysis. *BMC Psychiatry, 11*(1), 200. https://doi.org/10.1186/1471-244X-11-200

Rodebaugh, T. L., Heimberg, R. G., Schultz, L. T., & Blackmore, M. (2010). The moderated effects of video feedback for social anxiety disorder. *Journal of Anxiety Disorders, 24*(7), 663–671. https://doi.org/10.1016/j.janxdis.2010.04.007

Rouf, K. (2004). In J. Bennett-Levy, G. Butler, M. Fennell, A. Hackmann, M. Mueller, & D. Westbrook (Eds.), *Oxford guide to behavioural experiments in cognitive therapy*. Oxford University Press. https://doi.org/10.1093/med:psych/9780198529163.001.0001.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573–583. https://doi.org/10.1016/S0006-3223(02)01866-8

Schreiber, F., Heimlich, C., Schweitzer, C., & Stangier, U. (2015). Cognitive therapy for social anxiety disorder: The impact of the "Self-Focused Attention and Safety

Behaviours Experiment" on the course of treatment. *Behavioural and Cognitive Psychotherapy, 43*(2), 158–166. https://doi.org/10.1017/S1352465813000672

Shuman, V. (2013). Studying the social dimension of emotion regulation. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00922

Sosic, Z., Gieler, U., & Stangier, U. (2008). Screening for social phobia in medical in- and outpatients with the German version of the social phobia inventory (SPIN). *Journal of Anxiety Disorders, 22*(5), 849–859. https://doi.org/10.1016/j.janxdis.2007.08.011

Springer, K. S., Levy, H. C., & Tolin, D. F. (2018). Remission in CBT for adult anxiety disorders: A meta-analysis. *Clinical Psychology Review, 61*, 1–8. https://doi.org/10.1016/j.cpr.2018.03.002

Stangier, U., Clark, D. M., Ginzburg, D. M., & Ehlers, A. (2016). *Soziale angststörung* (2nd ed.). Hogrefe.

Stangier, U., & Heidenreich, T. (2005). Liebowitz soziale angst skala. In *Collegium internationale psychiatriae scalarum* (5th ed.). Hogrefe.

Stangier, U., Heidenreich, T., Peitz, M., Lauterbach, W., & Clark, D. M. (2003). Cognitive therapy for social phobia: Individual versus group treatment. *Behaviour Research and Therapy, 41*(9), 991–1007. https://doi.org/10.1016/S0005-7967(02)00176-6

Stangier, U., Schramm, E., Heidenreich, T., Berger, M., & Clark, D. M. (2011). Cognitive therapy vs interpersonal psychotherapy in social anxiety disorder: A randomized controlled trial. *Archives of General Psychiatry, 68*(7), 692. https://doi.org/10.1001/archgenpsychiatry.2011.67

WHO World Mental Health Survey Collaborators, Stein, D. J., Lim, C. C. W., Roest, A. M., De Jonge, P., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Benjet, C., Bromet, E. J., Bruffaerts, R., De Girolamo, G., Florescu, S., Gureje, O., Haro, J. M., Harris, M. G., He, Y., Hinkov, H., Horiguchi, I., … Scott, K. M. (2017). The cross-national epidemiology of social anxiety disorder: Data from the world mental health survey initiative. *BMC Medicine, 15*(1), 143. https://doi.org/10.1186/s12916-017-0889-2

Stenzel, N., Krumm, S., & Rief, W. (2010). Therapieplanung mithilfe des Interviews zur operationalisierten Fertigkeitsdiagnostik (OFD). *Verhaltenstherapie, 20*(2), 109–117. https://doi.org/10.1159/000293364

Trivedi, M. H., Rush, A. J., Ibrahim, H. M., Carmody, T. J., Biggs, M. M., Suppes, T., Crismon, M. L., Shores-Wilson, K., Toprac, M. G., Dennehy, E. B., Witte, B., & Kashner, T. M. (2004). The inventory of depressive symptomatology, clinician rating (IDS-C) and self-report (IDS-SR), and the quick inventory of depressive symptomatology, clinician rating (QIDS-C) and self-report (QIDS-SR) in public sector patients with mood disorders: A psychometric evaluation. *Psychological Medicine, 34*(1), 73–82. https://doi.org/10.1017/S0033291703001107

Von Glischinski, M., Willutzki, U., Stangier, U., Hiller, W., Hoyer, J., Leibing, E., Leichsenring, F., & Hirschfeld, G. (2018). Liebowitz social anxiety scale (LSAS): Optimal cut points for remission and response in a German sample. *Clinical Psychology & Psychotherapy, 25*(3), 465–473. https://doi.org/10.1002/cpp.2179

Wang, Y.-P., & Gorenstein, C. (2013). Psychometric properties of the beck depression Inventory-II: A comprehensive review. *Revista Brasileira de Psiquiatria, 35*(4), 416–431. https://doi.org/10.1590/1516-4446-2012-1048

Warnock-Parkes, E., Wild, J., Thew, G., Kerr, A., Grey, N., & Clark, D. M. (2022). 'I'm unlikeable, boring, weird, foolish, inferior, inadequate': How to address the persistent negative self-evaluations that are central to social anxiety disorder with cognitive therapy. *The Cognitive Behaviour Therapist, 15*, Article e56. https://doi.org/10.1017/S1754470X22000496

Weisman, J. S., & Rodebaugh, T. L. (2018). Exposure therapy augmentation: A review and extension of techniques informed by an inhibitory learning approach. *Clinical Psychology Review, 59*, 41–51. https://doi.org/10.1016/j.cpr.2017.10.010

Wells, A., Clark, D. M., Salkovskis, P., Ludgate, J., Hackmann, A., & Gelder, M. (2016). Social phobia: The role of In-Situation safety behaviors in maintaining anxiety and negative beliefs – Republished article. *Behavior Therapy, 47*(5), 669–674. https://doi.org/10.1016/j.beth.2016.08.010

Wells, A., White, J., & Carter, K. (1997). Attention training: Effects on anxiety and beliefs in panic and social phobia. *Clinical Psychology & Psychotherapy, 4*(4), 226–232. https://doi.org/10.1002/(SICI)1099-0879(199712)4:4<226::AID-CPP129>3.0.CO;2-M

Yoshinaga, N., Matsuki, S., Niitsu, T., Sato, Y., Tanaka, M., Ibuki, H., Takanashi, R., Ohshiro, K., Ohshima, F., Asano, K., Kobori, O., Yoshimura, K., Hirano, Y., Sawaguchi, K., Koshizaka, M., Hanaoka, H., Nakagawa, A., Nakazato, M., Iyo, M., & Shimizu, E. (2016). Cognitive behavioral therapy for patients with social anxiety disorder who remain symptomatic following antidepressant treatment: A randomized, assessor-blinded, controlled trial. *Psychotherapy and Psychosomatics, 85*(4), 208–217. https://doi.org/10.1159/000444221

Zaider, T. I., Heimberg, R. G., Fresco, D. M., Schneier, F. R., & Liebowitz, M. R. (2003). Evaluation of the clinical global impression scale among individuals with social anxiety disorder. *Psychological Medicine, 33*(4), 611–622. https://doi.org/10.1017/S0033291703007414