Cognitive bias modification and CBT as early interventions for adolescent social and test anxiety: Two-year follow-up of a randomized controlled trial

Eva de Hullu a, 1, 3, B. Esther Sportel b, a, 2, 3, Maaike H. Nauta a, b, *, Peter J. de Jong a

a Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands
b Department of Psychiatry, University Medical Center Groningen, University of Groningen P.O. Box 30.001, 9700 RB Groningen, The Netherlands

Abstract

Background and Objectives: This two-year follow-up study evaluated the long-term outcomes of two early interventions that aimed at reducing social and test anxiety in young adolescents at risk for developing social anxiety disorder.

Methods: In this RCT, moderately socially anxious adolescents (N=240, mean age 13.6 years) were randomly assigned to a 10-week internet-based multifaceted cognitive bias modification training (CBM), a 10-week school-based cognitive behavioral group training (CBT), or a no-intervention control condition. Using multiple imputation, this study examined the changes in primary and secondary outcome measures from pretest to follow-up in a repeated measures design.

Results: Primary outcome: Self-reported social and test anxiety generally decreased from pre-test to two-year follow-up, regardless of treatment condition. The percentage of adolescents who developed a social anxiety disorder was very low (6%) and similar across conditions. Secondary outcome: There were beneficial changes in self-esteem, self-reported prosocial behaviors, and fear of negative evaluation, but none of these were related to treatment condition. Automatic social-threat associations did not significantly change. The CBM intervention was effective in changing interpretative bias as indexed by the Recognition Task but this long-term effect did not transfer to the Adolescent Interpretation and Belief Questionnaire.

Limitations: There was a substantial (50%) though seemingly non-selective attrition at follow-up.

Conclusions: This RCT does not support the longer-term efficacy of school-based CBT or CBM as an early intervention for social and test anxiety. Rather, it emphasizes the positive ‘natural’ course of highly socially anxious adolescents over two years.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Social anxiety disorder (SAD) is the most common mental disorder in adolescence (Wittchen & Fehm, 2003) and is characterized by a fear of negative evaluation by others. Several common manifestations of social anxiety in adolescents (e.g., fear of reading aloud, taking tests, asking teachers questions), overlap with test anxiety (Bogels, Alden, Beidel, Clark, Pine, Stein et al., 2010), and test anxiety has been described as a component of social anxiety that is a major concern in school settings (Beidel and Turner, 2007). SAD is associated with poor social skills, reduced social interactions, low self-esteem, and low school performance (Stein, 2000). Test anxiety is not limited to social anxiety, interfering with social functioning, but also has an impact on academic performance. Higher levels of test anxiety are associated with impaired performance on tests, especially in high-stake tests (McDonald, 2010). Given this large impact on present and future functioning, early intervention seems of great importance.

http://dx.doi.org/10.1016/j.jbtep.2016.11.011
0005-7916/© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Recognizing the need expressed by schools for evidence-based interventions specifically targeting social and test anxiety in adolescents, our research group evaluated the efficacy of two early interventions in a sample of adolescents in the first years of secondary school with elevated levels of social and/or test anxiety (Sportel, de Hullu, de Jong, & Nauta, 2013). Based on a series of recent studies providing evidence that social anxiety can be reduced following successful modification of interpretive bias (Beard & Amir, 2008) or attentional bias (Amir, Beard, Taylor, Klumpp, Elias, Burns et al., 2009), we tested the efficacy of a 10-week internet-delivered Cognitive Bias Modification training (CBM) with a focus on modifying both attentional and interpretive bias as a cost and time-efficient strategy to reduce social and test anxiety. As a second strategy, we included a 10-week Cognitive Behavioral Group training (CBT) based on current golden-standard treatment protocols (Kendall, Hudson, Choudhury, Webb, & Pimentel, 2005; Warner, Fisher, Shront, Rathor, & Klein, 2007) adjusted for the purpose of early intervention in a Dutch sample of adolescents. CBT resulted in a decrease of test anxiety that was still evident at one-year follow-up, whereas the effect of CBM on test anxiety was small and not significant. In addition, CBT resulted in a stronger reduction of social anxiety symptoms than the control condition at six-months follow-up, with a moderate effect size, with a similar tendency for CBM, with a small effect size (Sportel et al., 2013). Unexpectedly, there was no significant difference between the active conditions relative to the no-treatment control group in reducing social anxiety symptoms at one-year follow-up.

However, to effectively test the long-term effect of a preventive early intervention, it is important to follow participants over a prolonged period of time. Therefore, the first aim of the present study was to test the efficacy of our CBT and CBM interventions in reducing social and text anxiety at a longer term (two-year) follow-up, adding to the previous study on the short-term gains (Sportel et al., 2013). In addition, we evaluated the efficacy of both interventions to prevent the development of social anxiety disorder, since the development of a disorder may warrant such a longer time horizon.

Furthermore, we evaluated the longer-term effects on a broader range of predefined secondary outcome measures. We chose outcome measures that may be similarly or differentially influenced by CBT and CBM. First, we tested the impact of our interventions on enhancing self-esteem. Low self-esteem has been identified as a risk factor for developing symptoms of anxiety and depression (Sowislo & Orth, 2013; van Tuijl, de Jong, Sportel, de Hullu, & Nauta, 2014) whereas high self-esteem has been argued to act as an anxiety buffer (Greenberg, Solomon, Pyszczynski, Rosenblatt, Burling, Lyon et al., 1992). Therefore, improvement of adolescents’ self-esteem might help prevent the further development of social and test anxiety. Both CBT and CBM may have a positive impact on self-esteem, given that our CBM contained an evaluative conditioning task aimed at enhancing implicit self-esteem, and CBT included the development of positive self-statements in cognitive restructuring. Moreover, the experience of exposure may give rise to a sense of mastery thereby enhancing self-esteem.

Second, we examined the influence of both interventions on social skills. It has been shown that social skills are generally impaired in individuals with social anxiety (Alfano, Beidel, & Turner, 2006; Beidel, Rao, Scharfstein, Wong, & Alfano, 2010), and social skills training is a common intervention for social anxiety in youth (Scharfstein & Beidel, 2011). In CBT, participants are encouraged to engage in approach behaviors in social situations during the exposure exercises. We, therefore, expected that CBT in particular would have a beneficial effect on prosocial behaviors.

Third, we included an implicit performance measure to examine whether the current interventions were effective in modifying participants’ automatic associations between social situations and indicators of a negative outcome (de Hullu, de Jong, Sportel, & Nauta, 2011). Attesting to the importance of automatic threat associations in the persistence of fear, it has been shown that residual automatic associations predict return of fear after exposure therapy for speech anxiety (Vasey, Harbaugh, Buffington, Jones, & Fazio, 2012). Given the focus of CBM on modifying more automatic information processing biases, we anticipated that especially CBM would have beneficial effects on the automatic threat associations.

Fourth, we tested the influence of CBT and CBM on adolescents’ fear of negative evaluation. Current cognitive models of social anxiety propose that social anxiety is derived in part from social-evaluative fears (e.g., Clark & Wells, 1995; Rapee & Heimberg, 1997). Because of the alleged role of perceived social-evaluative threats in the development of social anxiety it seems relevant to examine whether our preventive interventions may also affect adolescents’ fear of negative evaluation. Both CBT and CBM target negative cognitions about the self, so we expected positive effects of both interventions.

Finally, it has been argued that for an adequate evaluation of the relevance of CBM as a clinical intervention, it is important to determine whether CBM, the procedure that is intended to bring about change in cognitive bias and the process of actual cognitive bias change (e.g., MacLeod & Clarke, 2015). Only if the procedure is effective in modifying the cognitive bias, effects on anxiety vulnerability may be expected (e.g., Clarke, Notebaert, & MacLeod, 2014). Therefore, we also examined whether adolescents’ cognitive bias differentially changed from pre-test to two-year follow-up across the three conditions. Analysis of the immediate effects of the current CBM intervention supported the short term efficacy of our intervention as a means to modify interpretation bias, whereas the intervention failed to show robust effects on attentional bias (Sportel et al., 2013). We, therefore, restricted the follow-up analyses to the indices of interpretation bias.

In sum, the first aim of the current follow-up study was to test the longer-term impact of CBT and CBM on symptoms of test anxiety and social anxiety as indexed by self-reports and diagnoses of social anxiety disorder based on a clinical interview. As a second aim, the study examined the more general influence of both early interventions on self-esteem, (self-reported) social skills, automatic threat associations, and fear of negative evaluation.

2. Materials and methods

2.1. Design & ethics statement

The study used a multi-arm parallel group approach and employed a stratified design with balanced randomization (1:1:1). It was approved by the medical ethics committee of the University Medical Center Groningen. Participants, together with at least one parent or caretaker, provided written informed consent prior to the start of the study. The study was registered in the Dutch trial register with number NTR965 (http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=965). The trial can also be found by using either 965 (trial registration number) or PASTA (acronym referring to the Project on Adolescent Social and Test Anxiety) as a search term on www.trialregister.nl). Power analysis showed that for a medium effect, with a power of 0.80, within three groups, with an alpha of 0.05 (one-sided), the sample size had to be 52 for each condition. Because of anticipated drop-out we aimed at 75 participants per condition (total sample size of 225 adolescents). Recruitment took place in 2007 and 2008; all assessments took place between 2007 and 2012.
2.2. Participants

A total of 5318 adolescents in the first and second year of regular secondary schools were invited to participate in this study. Fig. 1 displays a flow diagram of this study.

Of the invited adolescents, 1811 participants were screened for social and test anxiety by means of the Revised Child Anxiety and Depression Scale (RCADS; Chorpita, Yim, Moffit, Umemoto, & Francis, 2000) and the Dutch version of the Spielberger’s Test Anxiety Inventory (TAI; Van der Ploeg, 1988). Participants scoring above cut-off for social and/or test anxiety (n = 516) were invited for a clinical assessment using the Anxiety Disorders Interview Schedule for Children (ADIS-C; Albano & Silverman, 1996). Cut-off scores for girls were >10 on RCADS social phobia and >43 on TAI; cut-off scores for boys were >9 on RCADS social anxiety and >38 on TAI. The RCADS cut-off scores were based on the 75th percentile in a large Dutch cohort of young adolescents (N = 2230, the TRAILS-study, Huisman, Oldhinkel, De Winter, Minderaa, De Bildt, Huizink et al., 2008), TAI cut-off scores were based on the 75th percentile in the Dutch manual (Van der Ploeg, 1988). Screening took place in two waves, including 11 schools in the first year and 13 schools in the second year. 

Based on the ADIS-C, adolescents with moderate-level social anxiety were included in the current study (N = 240; aged 12–16 years (90% aged 13–14 years; 66 boys). Moderate level of social anxiety was defined as having symptoms of SAD, and a Clinician Severity Rating of 4 or lower (CSR scale 0–8). For ethical reasons, adolescents with DSM-IV diagnoses other than anxiety and/or with severely interfering anxiety diagnoses (CSR > 5) and/or who expressed a need for regular treatment were referred to regular mental health centers. After the pre-test, participants were cluster randomized at school-level over one of three conditions and were stratified by the size of the school (see Fig. 1). Stratification per school size guaranteed that the number of participants would be similar across conditions. The allocation of the schools was done by the project leader, by blindly drawing same size papers with the conditions CBT, CBM or Control from a bowl (in the presence of the project leader) and thus would receive no training. It was stated that they were free to seek treatment if they felt the need, but none of the participants reported actually having received mental health treatment elsewhere during the intervention period of this study.

3. Interventions

The CBM intervention consisted of 20 sessions (40 min each), delivered twice a week via the internet. The backbone of CBM consisted of tasks to modify interpretation (9 sessions of 60 trials) and attention bias (8 sessions of 450 trials) (for details see Sportel et al., 2013). We also included a task (3 sessions of 500 trials) to strengthen the association between social-evaluative situations and positive outcomes, and a short evaluative conditioning task (Baccus, Baldwin, & Packer, 2004; Clerkin & Teachman, 2010) to enhance implicit self-esteem (240 trials; added to 10 CBM sessions). Each week, participants received an e-mail with links to two training sessions (Table 1), and were reminded if they did not complete a session.

The CBT intervention consisted of ten weekly sessions of 1.5 h that were delivered in groups (3–10 participants) by a licensed (CBT) psychologist, at school (PASTA training by Sportel and Nauta, 2007; for more details, see Sportel et al., 2013). Components were: Psycho-education (session 1 and 2); Task concentration training (session 3 and 4); Cognitive restructuring (session 5 and 6); Exposure (session 7, 8 and 9). The last session (10) focused on personal pitfalls and relapse prevention. Participants also received homework assignments.

On average, participants in the CBM condition completed 8.5 out of 20 CBM sessions (SD = 6.9). Participants in the CBT condition attended 6.7 sessions out of 10 sessions (SD = 3.3). A number of participants in the CBM condition (n = 16) did not start the training, mostly due to technical difficulties.

3.1. Control group

The adolescents in the schools that were randomly allocated to the control condition received no intervention. After the pre-test, they received a letter explaining that they formed the control group and thus would receive no training. It was stated that they were free to seek treatment if they felt the need, but none of the participants reported actually having received mental health treatment elsewhere during the intervention period of this study.

3.2. Measures

3.2.1. Process measures: changes in interpretative bias

Two tasks were used to assess changes in interpretation bias, namely the Recognition Task (Salemink & van den Hout, 2010) and the Adolescent Interpretation and Belief Questionnaire (AIBQ; Miers, Blöte, Bogels, & Westenberg, 2008). The scenarios in the Recognition Task were adapted for adolescents in a school setting. Participants read ten scenarios of ambiguous social situations on a computer screen, and then responded to a word puzzle containing a question related to the scenario, to check whether they read the information well. The scenarios remained ambiguous. After reading the ten scenarios, the title of the description was repeated and adolescents were asked to rate the similarity (1 = very similar in meaning to 4 = very different in meaning) of four different interpretations (positive, negative, neutral, or irrelevant). Scores were reversed, leading to higher scores reflecting a stronger interpretive bias. In the current study, we used a difference score (positive bias minus negative bias) to assess interpretative bias. In the AIBQ, adolescents also read ambiguous scenarios, including both social and non-social situations. Then, they were asked to read three different interpretations of the situation, namely neutral, positive or negative, and rate whether this explanation also popped into their minds (scale 1–5), with lower scores indicating a lower bias.

3.2.2. Primary outcome measures

Social anxiety symptoms were indexed by the social phobia subscale (9 items) of the Revised Child Anxiety and Depression Scale (RCADS; Chorpita et al., 2000) with items rated on a 4-point scale ranging from 0 (never) to 3 (always). Internal consistency was satisfactory (at pre-test α = 0.79).

Test anxiety was indexed by the Spielberger Test Anxiety Inventory (Spielberger TAI; Van der Ploeg, 1988), with 20 items rated on a 4-point scale, ranging from 1 (almost never) to 4 (all the time). Reliability proved to be excellent (at pre-test α = 0.95).

To assess the presence of SAD at pre-test and at two-year follow-up, we carried out clinical interviews using the anxiety and mood sections of the ADIS-C (Albano & Silverman, 1996). We focused on the social phobia diagnosis (present or not, rated as 1/0). A social phobia diagnosis was established if youth met all relevant criteria including a Clinician Severity Rating of 4 or higher (scale 0–8). The interrater-reliability for a diagnosis of social phobia was very high with 99.7% overlap (based on ratings by a psychologist and independent rater scoring a random selection [n = 89] of the available ADIS-C audio-taped interviews [n = 248] from pre-test assessment).
3.2.3. Secondary outcome measures

Self-esteem was measured with a Dutch adaptation of the RSES (Mayer, Muris, Meesters, & Zimmermann-van Beuningen, 2009) which consists of fifteen items based on the original RSES (Rosenberg, 1989) that were rated on a five-point scale from 0 (completely not true) to 4 (completely true). The internal consistency of the RSES was excellent (at pretreatment $\alpha = 0.90$).

Prosocial behaviors were assessed with the “Vragenlijst Sociale Vaardigheden van Jongeren” (VSVJ; Hulstijn, Cohen-Kettenis, Mellenbergh, Boomsma, Blonk, Prins et al., 2006) which is derived from

---

**Table 1**
Order of tasks in the CBM training.

<table>
<thead>
<tr>
<th>Week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>First task</td>
<td>IB</td>
<td>ABA</td>
<td>IB</td>
<td>AA</td>
<td>ABB</td>
<td>IB</td>
<td>IB</td>
<td>AA</td>
<td>ABB</td>
<td>SE</td>
</tr>
<tr>
<td>Second task</td>
<td>IB</td>
<td>ABA</td>
<td>AA</td>
<td>IB</td>
<td>ABB</td>
<td>ABA</td>
<td>ABB</td>
<td>IB</td>
<td>IB</td>
<td>SE</td>
</tr>
<tr>
<td>SE</td>
<td>SE</td>
<td>SE</td>
<td>SE</td>
<td>SE</td>
<td>SE</td>
<td>SE</td>
<td>SE</td>
<td>SE</td>
<td>SE</td>
<td>SE</td>
</tr>
</tbody>
</table>

Note. IB = interpretive bias; ABA = attentional bias, first version; ABB = attentional bias, second version; AA = automatic associations; SE = implicit self-esteem.
the Matson Evaluation of Social Skills with Youngsters (MESSY; Matson, Neal, Fodstad, Hess, Mahan, Rivet et al., 2010). Like the MESSY, the VSVJ has two subscales; prosocial behaviors (16 items) and inappropriate behaviors (12 items). For the current study, we only used the prosocial behaviors subscale (at pretest $\alpha = 0.86$).

Threat-related automatic associations were assessed by means of a Single Target Implicit Association Test (stIAT) with the target category ‘social or school activity’, and attribute labels positive versus negative outcome (see de Hullu et al., 2011 for details). StIAT scores were computed according to the algorithm proposed by Greenwald, Nosek and Banaji (2003), which has shown to perform also best in a laboratory setting (Glashouwer, Smulders, De Jong, Roefs, & Wiers, 2013). A high D-score indicates relatively strong automatic associations between social or school activities and positive outcomes. Split-half reliability as indexed by Spearman-Brown corrected coefficient was 0.72. To ensure the validity of this measure, the stIAT data were discarded if for more than 10% of the trials participants’ response times were faster than 300 ms, or for more than 1% of the trials slower than 10,000 ms, or if participants’ error percentage was higher than 2SD above the mean (van Tuyl et al., 2014).

Fear of negative evaluation was assessed using the 12-item Brief Fear of Negative Evaluation-II (BFNE-II; Carleton, McCreary, Norton, & Asmundson, 2006; BFNE—II—NL van Wees-Geraad & de Jong, 2007) with 5-point Likert scale ranging from 0 (does not fit me at all) to 4 (fits me very well). Internal consistency was excellent ($\alpha = 0.95$).

3.3. Procedure

The assessments took place at school. Tasks and questionnaires were presented in fixed order. After the pre-test, participants were informed about the assigned condition. Post-test was after 12 weeks, followed by assessments at six months, one-year, and two-year follow-up. Participants received a gift certificate (€5) for each assessment. The ADIS-C at two-year follow-up was conducted via telephone. Assessors of questionnaires and interviews remained blind for participants’ condition.

3.4. Statistical analyses

Analyses were conducted on the basis of intent-to-treat, so on the basis of all participants that were randomized. As a second step, sensitivity analyses were conducted to check if the effects were similar in the group of adolescents who followed at least half of the CBM or CBT sessions that were offered.

We used multiple imputation using SPSS 23 software to impute the missing data at the two-year follow-up. Multiple imputation is considered a good strategy for dealing with missing data of incomplete cases, especially when the missingness may not be completely at random (e.g., van Buuren, 2011). Multiple imputation has three phases: 1. Imputing the data in $m$ datasets. 2. Analyzing the data in each data set, and 3. Pooling the results of all data sets to come to inference. To come to a good estimation, we selected the following predictors to estimate the value of the variables at two-year follow-up: age, gender, baseline data of all outcome measures (RCADS-SAD, TAI, the clinical severity rating of the ADIS-C, RSES, MESSY, stIAT, FNE), as well as two process measures of interpretation bias, and available data of these outcome and process measures at post-treatment, 6 and 12 months follow-up. Thus, all available data at various assessment points was weighed into the estimation of the missing scores at two-year follow-up. Predictive mean matching was used as a strategy to impute the missing data in 40 different datasets.

For each outcome variable, 3 (condition) by 2 (time) repeated measure ANOVAs were conducted on each dataset separately. Since SPSS does not provide the F-statistics on the pooled data, we used the combination rules as described by van Ginkel and Kroonenberg (2014), using a SPSS macro (Van Ginkel, 2014). This procedure provides the possibility of pooling the results of all separate data- sets into pooled F-statistics. Following this procedure, overall F-statistics were computed (full model, intercept, time, condition, and time*condition). Main effects for treatment condition were not hypothesized and were only reported in the Results section if significant. If the interaction F-test was significant, parameter estimates of the fixed effects were tested with a t-statistic to examine the differential effects of time*CBT (vs. control) and time*CBM (vs. control). The fraction of missing information (fmi) is provided for each parameter estimate in the model. Fmi reflects the percentage of sampling variance that is attributable to missing data and provides an indication of the severity of the missing data problem (with higher scores reflecting higher severity).

For each outcome measure and each treatment condition, Cohen’s $d$ effect size was calculated for the pre-test to two-year follow-up period by dividing the difference in means by the pooled standard deviation, based on the imputed datasets but taking into account the observed N at two-year follow-up.

4. Results

4.1. Descriptive statistics of missing data

In total, 121 of 240 participants (50%) completed the two-year follow-up assessment: 40 (46.5%) of the 86 participants in the CBM condition, 46 (54.8%) of the 84 participants in the CBT condition, and 35 (50.0%) of the 70 participants in the control condition. See Fig. 1 for a complete overview of participation in measurements. To check whether missingness at follow-up was related to pre-test scores of the participants, we first compared the pre-test scores of those who did participate in the two-year follow-up with those who did not. These tests showed that the groups did not differ at pre-test (ADIS-C SAD: $X^2(1) = 0.06, p = 0.85$; RCADS-SAD: $t = 0.11, p = 0.91$; TAI: $t = 0.99, p = 0.32$).

Second, we tested whether at pre-test the scores of participants who participated in the follow-up assessment varied across conditions. No differences were found between the assessment completers and assessment non-completers across the three conditions at pre-test (ADIS-C SAD: $X^2(2) = 3.62, p = 0.16$; RCADS-SAD: $F(2,118) = 0.21, p = 0.82$; TAI: $F(2,2118) = 0.15, p = 0.86$).

In conclusion, missing data analyses did not reveal a difference between assessment completers and assessment non-completers with regard to the outcome measures in the whole group or within the treatment conditions.

4.2. Descriptive statistics

The means and standard deviations for all primary and secondary outcome measures at pre-test and two-year follow-up are presented in Table 2. At pre-test, all conditions showed similar scores on the primary outcome measures (RCADS-SAD $F(2,239) = 0.30, p = 0.74$; Spielberger TAI $F(2,239) = 0.06, p = 0.94$; ADIS-C SAD $X^2(2) = 1.36, p = 0.54$).

4.2.1. Process outcomes: effects of CBT and CBM on interpretation bias

Table 3 presents the means, standard deviations of the pre-test and follow-up assessments of the interpretation bias (process) measures (AIBQ and RT). The AIBQ scores showed a significant overall decrease from pre-test to two-year follow-up $F(1,1197) = 61.2, p < 0.001$. There was, however, no significant
interaction effect $F(2,158.4) = 1.02, p = 0.36$. Thus the different conditions did not lead to different changes in interpretation bias as measured with the ALIB. For the Recognition Task (RT), there was also a significant change in interpretation bias scores from pre-test to two-year follow-up $F(1,150.2) = 71.2, p < 0.001$. Importantly, the decrease in interpretation bias as indexed by the RT varied across conditions ($F(2,162.8) = 6.81, p < 0.05$. Post hoc t-tests of the parameter estimates, comparing the effect in each active condition versus the control group, revealed that the critical interaction effect was significant for CBM ($t(69) = -3.41, p = 0.001$) but not for CBT ($t(76) = 0.63, p = 0.53$), meaning that adolescents in the CBM condition reported larger changes on the RT than youth in the control condition.

5. Effects of CBT and CBM

5.1. Primary outcome measures

5.1.1. Social anxiety (RCADS-SAD)

There was a significant overall decrease of RCADS-SAD scores from pretest to two-year follow-up $F(1,104.0) = 58.50, p < 0.001$. The active treatment conditions did not significantly add to the decrease of social anxiety over time, as evidenced by a non-significant interaction effect, $F(2,166.2) = 0.45, p = 0.64$.

5.1.2. Test anxiety (Spielberger TAI)

There was a significant overall decrease of TAI scores from pretest to two-year follow-up $F(1,110.7) = 52.1, p < 0.001$. The interaction effect was not significant $F(2,179.6) = 1.14, p = 0.32$, indicating that the active conditions did not lead to a relatively strong reduction in test anxiety.

5.1.3. Social anxiety disorder (ADIS-C)

At two-year follow-up, the number of social anxiety disorder diagnoses in the observed data was 5.9% (7 of 119 adolescents), with 1 out of 45 (2.2%) in the CBT condition, 4 out of 42 (9.5%) in the CBM condition, and 2 out of 32 (6.3%) in the control condition. The pattern of results was similar in the imputed datasets, with pooled estimates of social anxiety disorder being 4 of 84 youths in the CBT condition (4.8%), 8 of 86 in the CBM condition (9.3%), and 3 of 70 youth in the control condition (4.3%). Since only a very small fraction of the participants received or were estimated to receive a diagnosis of SAD, these data could not be meaningfully subjected to statistical analysis.

5.2. Secondary outcome measures

5.2.1. Self-esteem (RSES)

There was an overall increase of self-esteem between pre-test and follow-up, $F(1,128.1) = 34.4, p < 0.001$, but there was no additional benefit for the active conditions, as evidenced by the absence of a significant interaction effect $F(2,161.6) = 0.93, p = 0.39$.

5.2.2. Prosocial behavior (VSVJ/MESSY)

There was a main effect for time, indicating that there was an overall increase of prosocial behaviors over time, $F(1,128.1) = 34.4, p < 0.001$. The active conditions did not significantly add to treatment gains as evidenced by a nonsignificant interaction effect, $F(2,166.4) = 1.45, p = 0.24$.

5.2.3. Implicit social-threat associations (stiAT)

The model showed no significant main effect of Time $F(1,99.7) = 1.33, p = 0.25$, indicating that, overall, the strength of automatic threat associations remained stable over time. This pattern was similar for all three conditions, as was evidenced by the absence of a significant interaction between time and condition, $F(2,163.8) = 1.21, p = 0.30$.

5.2.4. Fear of negative evaluation (BFNE-II)

The model showed that, overall, the BFNE scores declined over

---

Table 2

Descriptives of all measures at pre-test and at two-year follow-up for the three treatment conditions.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Pre-test</th>
<th>Two-year follow-up (observed data)</th>
<th>Two-year follow-up (estimated means)</th>
<th>Effect size $d$</th>
<th>95% confidence interval $d$ (lower – upper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social phobia (RCADS-SAD)</td>
<td>CONTR</td>
<td>13.27 (4.52)</td>
<td>10.83 (5.38)</td>
<td>0.63</td>
<td>0.21 – 1.04</td>
</tr>
<tr>
<td></td>
<td>CBT</td>
<td>13.11 (4.26)</td>
<td>8.96 (5.16)</td>
<td>0.86</td>
<td>0.48 – 1.23</td>
</tr>
<tr>
<td></td>
<td>CBM</td>
<td>13.64 (4.95)</td>
<td>9.53 (5.87)</td>
<td>0.79</td>
<td>0.40 – 1.18</td>
</tr>
<tr>
<td>Test anxiety (TAI)</td>
<td>CONTR</td>
<td>41.59 (13.23)</td>
<td>36.57 (12.88)</td>
<td>0.46</td>
<td>0.05 – 0.87</td>
</tr>
<tr>
<td></td>
<td>CBT</td>
<td>41.82 (13.28)</td>
<td>31.04 (10.10)</td>
<td>0.82</td>
<td>0.45 – 1.19</td>
</tr>
<tr>
<td></td>
<td>CBM</td>
<td>41.09 (13.94)</td>
<td>32.35 (12.12)</td>
<td>0.65</td>
<td>0.26 – 1.03</td>
</tr>
<tr>
<td>Self-esteem (RSES)</td>
<td>CONTR</td>
<td>30.23 (9.64)</td>
<td>34.74 (9.93)</td>
<td>-0.54</td>
<td>-0.95 – -0.12</td>
</tr>
<tr>
<td></td>
<td>CBT</td>
<td>30.02 (9.36)</td>
<td>37.57 (10.36)</td>
<td>-0.67</td>
<td>-1.03 – -0.29</td>
</tr>
<tr>
<td></td>
<td>CBM</td>
<td>29.06 (10.45)</td>
<td>38.33 (11.73)</td>
<td>-0.79</td>
<td>-1.17 – -0.40</td>
</tr>
<tr>
<td>Prosocial behavior (VSVJ/MESSY)</td>
<td>CONTR</td>
<td>56.77 (9.10)</td>
<td>59.17 (8.48)</td>
<td>-0.57</td>
<td>-0.98 – -0.15</td>
</tr>
<tr>
<td></td>
<td>CBT</td>
<td>56.42 (8.63)</td>
<td>61.13 (8.67)</td>
<td>-0.57</td>
<td>-0.93 – -0.20</td>
</tr>
<tr>
<td></td>
<td>CBM</td>
<td>57.02 (9.16)</td>
<td>61.55 (9.96)</td>
<td>-0.63</td>
<td>-1.01 – -0.25</td>
</tr>
<tr>
<td>Implicit associations (stiAT)</td>
<td>CONTR</td>
<td>-0.003 (0.28)</td>
<td>0.019 (0.27)</td>
<td>-0.11</td>
<td>-0.54 – 0.32</td>
</tr>
<tr>
<td></td>
<td>CBT</td>
<td>-0.027 (0.29)</td>
<td>-0.082 (0.26)</td>
<td>0.07</td>
<td>-0.31 – 0.34</td>
</tr>
<tr>
<td></td>
<td>CBM</td>
<td>-0.026 (0.33)</td>
<td>0.081 (0.28)</td>
<td>-0.30</td>
<td>-0.69 – 0.10</td>
</tr>
<tr>
<td>Fear of negative evaluation (BFNE-II)</td>
<td>CONTR</td>
<td>22.31 (11.13)</td>
<td>20.14 (11.54)</td>
<td>0.34</td>
<td>-0.07 – 0.74</td>
</tr>
<tr>
<td></td>
<td>CBT</td>
<td>22.40 (10.36)</td>
<td>17.35 (12.07)</td>
<td>0.49</td>
<td>0.13 – 0.86</td>
</tr>
<tr>
<td></td>
<td>CBM</td>
<td>23.52 (11.64)</td>
<td>17.47 (13.67)</td>
<td>0.49</td>
<td>0.11 – 0.87</td>
</tr>
</tbody>
</table>

* The means (SDs) of the stiAT effects differ from our previous article about the short term effects of this study because in the current analyses we used more stringent criteria to ensure the validity of the stiAT data and discarded stiAT effects if participants’ response times were faster than 300 ms for more than 10% of the trials, or slower than 10,000 ms in more than 1% of the trials, or if participants’ error percentage was higher than 2 SDs above the mean (cf. van Tuijl et al., 2014).
In conclusion, participants in the CBM condition showed a benefit in at least two of the measures ANOVA’s was on average 33% for main effects and interaction effects across primary and secondary outcomes (range 0.24–0.38). This means that the dataset benefitted from the multiple imputation procedure: after imputation, on average 33% of the sampling variance was attributable to missing data, while the real percentage of missing data is 50%.

5.2.5. Fraction of missing information

The fraction of missing information (fmi) in the repeated measures ANOVA’s was on average 0.33 for main effects and interaction effects across primary and secondary outcomes (range 0.24–0.38). This means that the dataset benefitted from the multiple imputation procedure: after imputation, on average 33% of the sampling variance was attributable to missing data, while the real percentage of missing data is 50%.

5.2.6. Sensitivity analyses

Sensitivity analyses were conducted for participants who received a minimum dosage of half of the sessions of the interventions. In the CBT condition, 62 of 84 adolescents completed at least 5 of 10 group sessions (74%), while in the CBM condition, 34 of 86 adolescents (40%) completed at least half of the sessions. All analyses were rerun in this subset of youth, together with the control group for which we included only those who had participated in at least two of five assessments (57 of 70 participants, 81%). Estimations of the frequency of social anxiety disorder (ADIS-C) at two-year follow-up were also low (11 of 153 adolescents, 7%), with the following distribution across the treatment conditions: in the control condition 5 of 57 participants (8.8%), in the CBT condition 4 of 62 (6.5%), and in the CBM condition 2 of 34 participants (5.7%). The analyses of the primary and secondary outcome measures yielded the same pattern of results as in the sample including the non-completers. can be requested from the authors and can be found in the online data repository.

6. Discussion

The major findings with regard to the primary outcome measures are as follows: First, social and test anxiety generally decreased from pre-test to two-year follow-up. Second, the active interventions did not add to a decrease in social or test anxiety. Thus, the favorable effects of CBT on social and test anxiety at shorter follow-up were no longer evident at two-year follow up. Third, the incidence of social anxiety disorder at two-year follow-up was very low and similar for all groups. With regard to the secondary outcome measures, there was a long-term overall increase in self-esteem and prosocial behaviors as well as a decrease of fear of negative evaluation. There was no overall change in automatic threat-related associations. For neither intervention group the pattern of these measures was significantly different from that of the control group.

For CBT, the effect sizes of the long-term changes in primary outcome measures fell in the range of large effects ($d = 0.86$ and 0.82 for social anxiety and test anxiety, respectively), and were in line with the effect sizes found in earlier studies on school-based interventions (Mychailyszyn, Brodman, Read, & Kendall, 2012). For CBM the effect sizes fell in the range between moderate and large effects ($d = 0.79$ and 0.65), whereas the changes of the primary outcome measures in the control condition were of moderate effect size ($d = 0.63$ and 0.46). The strong reduction of social and test anxiety in the control group together with the very low proportion of participants in this group who fulfilled the criteria of a social anxiety disorder at two-year follow-up (6%) clearly reduced the sensitivity of the current study to find surplus effects of the active interventions.

Although previous intervention studies (Sheffield, Spence, Rapee, Kowalenko, Wagnall, Davis et al., 2006) also showed an overall drop of anxiety scores in at risk groups, the decrease of anxiety within control conditions is typically in the small range (e.g., Mychailyszyn et al., 2012; for school-based interventions for anxiety). It is unclear why the decline of social anxiety in the current control condition was relatively large (i.e., in the medium range). One explanation may be that the participating schools all had a relatively good social climate since they were willing to participate in the current trial, or that the explanation of the study to the children and the frequent assessments already made a difference.

The failure to find a surplus effect of the current CBM procedure may also be due to a failure to actually modify adolescents’ cognitive bias. Although the CBM intervention showed a robust beneficial effect on interpretation bias as indexed with the task that was most similar to the modification procedure (i.e., the recognition task), an equivalent beneficial effect was absent when interpretation bias was measured with an instrument that was more divergent from the tasks used in the CBM intervention (i.e., Adolescent Interpretation and Belief Questionnaire). Perhaps, then, the long-term influence of the current CBM procedure on cognitive bias was not sufficiently strong to have a meaningful impact on adolescents’ vulnerability for social and test anxiety. This may also help explain why the short-term beneficial effect on automatic threat associations was no longer evident at two-year follow-up.

We selected adolescents with heightened levels of social and test anxiety, under the assumption that they were at risk for developing anxiety disorders. Contrary to our expectations of deteriorating in the no-intervention group, many adolescents improved on our primary and most of the secondary measures, regardless of receiving an intervention. On the basis of this finding one may argue that the prospects of moderately socially anxious adolescents are good enough over a longer period of time and do
not require particular interventions; it may be more efficient to treat anxiety disorders in mental health services once they have developed.

Some limitations of our study should also be mentioned. First, we chose a cluster-randomized design, with randomization at school level rather than at the individual level. This was preferred to prevent potential “contamination” across conditions which might occur when more than one condition would be assigned to a particular school. In addition, this procedure enhanced the feasibility of the project and guaranteed sufficient participants at each site for the CBT group intervention. However, the fact that we randomized at school level was not taken into account in the power analyses. It cannot be ruled out that the study was therefore underpowered. Second, it should be acknowledged that a large proportion of the participants (50%) missed one or more of the post-intervention assessments. Although pre-test scores were similar for participants who did and those who did not complete all assessments, it cannot be ruled out that there was selective attrition during the follow-up assessments. We did use multiple imputation and the fraction of missing information was thus reduced (ranging from 24 to 38% with a mean of 33%). Third, we excluded adolescents with highly interfering social anxiety disorders. This might have limited the sensitivity of finding treatment gains in the current study. However, the current selection procedure reflects the procedures in many high schools, where schools may offer early interventions to youths with moderate problems, whereas youth with disorders are referred to mental health services. Finally, it is important to emphasize that the current ingredients of the CBT intervention reflect the first generation of CBT interventions. Furthermore, the current CBT combined several components that were designed to address separate cognitive biases. For optimizing CBT it would be important for future research to investigate which components are most effective in modifying the various pertinent cognitive biases (Lau, 2015; Macleod et al., 2015). Thus, although the current failure to show a reliable effect on the primary outcome measures is in line with recent meta-analyses questioning the efficacy of CBT-procedures as a clinical tool (Cristea, Kok, & Cuypers, 2015; Cristea, Mogoça, David, & Cuypers, 2015), it might be premature to conclude that CBT can be discarded as a potentially effective clinical intervention.

7. Conclusion

The findings showed that the level of social and test anxiety in moderately anxious adolescents generally improved over a two-year time period. The results do not support the use of current CBT procedures as a stand-alone intervention for social and test anxiety, since we found neither short-term (Sporkel et al., 2013) nor long-term benefits in terms of a decrease in symptomatology. The CBT group interventions in schools did provide beneficial effects at 6- and 12-months follow-up (Sporkel et al., 2013), but no additional benefits over a two-year period. Schools need to consider whether the investment in such a training weighs up to the moderate short-term benefits and the lack of robust long-term effects.

Declaration of interest and role of funding organization

The authors declare no conflict of interest. Sporkel and Nauta have developed the CBT group training but have no financial interest in the manual. This research was supported by a grant from the Netherlands Organization for Health Research and Development, ZonMw, nr. 62200027. The funding organization was not involved in the collection, analysis and interpretation of data, nor in the writing of the report or in the decision to submit the article for publication.

Acknowledgement

The authors kindly acknowledge dr. van Ginkel of Leiden University in the Netherlands for offering additional advice to his publicly available SPSS macro on multiple imputation.

References


