Exposure to virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial

Kampmann, I.L.; Emmelkamp, P.M.G.; Hartanto, D.; Brinkman, W.P.; Zijlstra, B.J.H.; Morina, N.

DOI
10.1016/j.brat.2015.12.016

Publication date
2016

Document Version
Final published version

Published in
Behaviour Research and Therapy

Behaviour Research and Therapy 77 (2016) 147–156

Contents lists available at ScienceDirect

Behaviour Research and Therapy

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

Exposure to virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial

Isabel L. Kampmann [a, *], Paul M.G. Emmelkamp [c, d], Dwi Hartanto [b], Willem-Paul Brinkman [b], Bonne J.H. Zijlstra [e], Nexhmedin Morina [a]

[a] Department of Clinical Psychology, University of Amsterdam, Weesperplein 4, 1018 XA, Amsterdam, The Netherlands
[b] Interactive Intelligence Group, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands
[c] Netherlands Institute for Advanced Study, Meijboomlaan 1, 2242 PR, Wassenaar, The Netherlands
[d] King Abdulaziz University, Abdullah Sulayman, 22254, Jeddah, Saudi Arabia
[e] Department of Educational Sciences, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 4 September 2015
Received in revised form
19 December 2015
Accepted 23 December 2015
Available online 29 December 2015

Keywords:
Virtual reality
Exposure therapy
Social anxiety disorder
Social phobia
Social interaction

ABSTRACT

This randomized controlled trial investigated the efficacy of a stand-alone virtual reality exposure intervention comprising verbal interaction with virtual humans to target heterogeneous social fears in participants with social anxiety disorder. Sixty participants (M_{age} = 36.9 years; 63.3% women) diagnosed with social anxiety disorder were randomly assigned to individual virtual reality exposure therapy (VRET), individual in vivo exposure therapy (iVET), or waiting-list. Multilevel regression analyses revealed that both treatment groups improved from pre-to postassessment on social anxiety symptoms, speech duration, perceived stress, and avoidant personality disorder related beliefs when compared to the waiting-list. Participants receiving iVET, but not VRET, improved on fear of negative evaluation, speech performance, general anxiety, depression, and quality of life relative to those on waiting-list. The iVET condition was further superior to the VRET condition regarding decreases in social anxiety symptoms at post- and follow-up assessments, and avoidant personality disorder related beliefs at follow-up. At follow-up, all improvements were significant for iVET. For VRET, only the effect for perceived stress was significant. VRET containing extensive verbal interaction without any cognitive components can effectively reduce complaints of generalized social anxiety disorder. Future technological and psychological improvements of virtual social interactions might further enhance the efficacy of VRET for social anxiety disorder.

Social anxiety disorder (SAD) is defined as the fear of one or more social situations in which one might behave embarrassingly and be negatively evaluated by others (DSM-V; American Psychiatric Association, 2013). SAD is one of the most common mental disorders in the US population, with an estimated lifetime prevalence of 12.1% (Ruscio et al., 2008). Individuals who suffer from SAD can experience a reduced quality of life and significant impairments in various areas of functioning, such as work and interpersonal relationships (Wittchen, Fuetsch, Sonntag, Müller, & Liebowitz, 2000). However, only about one third of individuals with SAD seek treatment (Ruscio et al., 2008).

The most researched treatment for SAD is cognitive behavior therapy (CBT). CBT aims at modifying maladaptive cognitions and behavior using both cognitive (e.g., cognitive restructuring) and behavioural (e.g., exposure) strategies (Hofmann & Smits, 2008; Mayo-Wilson et al., 2014). During exposure therapy, participants encounter feared stimuli in situations containing social interaction until anxiety decreases and/or anxiety-related expectancies are violated. Traditional exposure exercises are usually practiced during therapy and as homework assignments. Interestingly, a meta-analysis of treatment efficacy found exposure therapy alone to be comparable to cognitive therapy and that the combination of both was no more effective than either one delivered exclusively (Powers, Sigmarsson, & Emmelkamp, 2008).

A relatively new form of exposure therapy is Virtual Reality Exposure Therapy (VRET). During VRET, participants are confronted with computer-generated stimuli (e.g. virtual social interaction) that can elicit elevated subjective levels of social anxiety (Morina,

* Corresponding author.
E-mail address: i.l.kampmann@uva.nl (I.L. Kampmann).

http://dx.doi.org/10.1016/j.brat.2015.12.016
0005-7967/

Brinkman, Hartanto, & Emmelkamp, 2014; Powers et al., 2013). Cumulative research suggests that VRET is effective in the treatment of several anxiety disorders (Meyerbröker & Emmelkamp, 2010; Morina, Ijntema, Meyerbröker, & Emmelkamp, 2015; Opriş et al., 2012; Parsons & Rizzo, 2008).

While VRET has been extensively studied in specific phobias, research on the efficacy of VRET in the treatment of SAD is still limited. Several studies suggest that VRET can reduce SAD symptoms (Anderson, Rothbaum, & Hodges, 2003; Anderson, Zimand, Hodges, & Rothbaum, 2005; Klinger et al., 2005). However, only three randomized controlled trials on the efficacy of VRET in SAD have been conducted (Anderson et al., 2013; Bouchard et al., 2015; Wallach, Safir, & Bar-Zvi, 2009). In the study by Wallach et al. (2009), VRET for public speaking anxiety, a specific social anxiety complaint, was combined with CBT and compared to CBT plus imagery exposure, and waiting-list. Results revealed that VRET plus CBT was effective in treating public speaking anxiety compared to waiting-list and as effective as CBT plus imagery exposure. However, participants in this study were not screened for a clinical diagnosis of SAD. Anderson et al. (2013) included participants with a SAD diagnosis and compared the efficacy of CBT plus VRET with CBT plus group exposure therapy. The authors reported that CBT plus VRET was as effective as CBT plus group exposure therapy. Nonetheless, the implications of the results of this study are rather limited by the inclusion of participants who had reported public speaking anxiety as their primary complaint and by the two different formats of treatment (i.e., individual vs. group).

In both the above trials, exposure exercises solely targeted public speaking-related anxiety and included only limited verbal interaction (i.e., answering questions). However, although fear of public speaking is the most common subtype of SAD, the majority of individuals with SAD report more than one fear (Ruscio et al., 2008), emphasizing the need for research on VRET targeting heterogeneous social fears. Moreover, a large number of feared social situations reported by individuals with SAD (e.g., talking to strangers or speaking up in a meeting) contain verbal interaction (Ruscio et al., 2008). As a consequence, incorporating extensive dialogues into VRET and thus going beyond answering a limited number of questions might improve the efficacy of VRET for SAD. In contrast to Anderson et al. (2013) and Wallach et al. (2009), Bouchard et al. (2015) included virtual scenarios in VRET targeting several social fears. They found individual CBT plus VRET to be effective compared to waiting-list and more effective than CBT plus in vivo exposure. However, all three studies investigated VRET in combination with CBT. Therefore, no conclusions can be drawn regarding the efficacy of VRET as stand-alone treatment and the possibility cannot be ruled out that the effects found were caused by CBT rather than VRET.

In summary, previous research on VRET is limited by investigating VRET only in combination with CBT, focussing mainly on fear of public speaking and including only limited verbal interaction. The incorporation of diverse virtual scenarios with social interaction that resembles real life interaction into VRET might more adequately target the idiosyncratic fears of participants with SAD. The aim of the present study was to single out the effects of pure VRET without any cognitive components and to adapt VRET to individuals with heterogeneous social fears by simulating social verbal interaction in a variety of virtual social situations believed to be relevant for treating individuals with SAD. In a randomized controlled trial, we examined the efficacy of VRET and in vivo exposure therapy (iVET) for adults with SAD and heterogeneous social fears. These active treatments were compared to a waiting-list control group. Both active treatments were administered in an individual format and were exposure-based only. It was hypothesized that relative to individuals in the waiting-list control group, participants in active conditions would report fewer social anxiety symptoms and would perform better on a behavioural assessment task at postassessment. Treatment gains were expected to be comparable for VRET and iVET at postassessment and 3-month follow-up.

1. Method

1.1. Participants

Participants were recruited via online and newspaper advertisements, the website of the ambulatory of the University of Amsterdam, and the project's website. Sixty participants (M_{age} = 36.9 years, age range: 18–65 years) meeting the criteria for a primary diagnosis of SAD according to the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2000) were included and randomly assigned to one of three conditions (20 participants each; see Fig. 1 for an overview of the randomization procedure and Table 1 for sample characteristics per condition). Exclusion criteria were a) psychotherapy for SAD in the past year; b) current use of tranquilizers or change in dosage of antidepressants in the past 6 weeks; c) a history of psychosis, current suicidal intentions, or current substance dependence; e) severe cognitive impairment; or f) insufficient command of the Dutch language. The average number of completed sessions was 8.50 (SD = 2.63) for VRET and 8.55 (SD = 2.68) for iVET. All participants received free treatment and a small monetary reward (22 Euro) for the completion of the follow-up assessment.

1.2. Measures

1.2.1. Screening and diagnostic measures

The Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998) was used for screening purposes before the in-person interview. The SIAS consists of 20 items assessing cognitive, affective, and behavioural responses to social interactions on a 5-point Likert scale. The SIAS possesses a high internal consistency and test-retest reliability (Cronbach's α = .93 and r = 0.92 respectively; Mattick & Clarke, 1998). Individuals scoring ≥29 were invited for an in-person diagnostic interview with a psychologist. We choose a slightly lower cut-off than reported in previous research to prevent false-negatives in this early stage of screening where the in-person intake was still to come (Heimberg, Mueller, Holt, Hope, & Liebowitz, 1992).

To assess the diagnosis of SAD and potential comorbidity, the Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID-I; First, Spitzer, Gibbon, & Williams, 1994) was administered prior to inclusion. All assessors were psychologists with a master degree in clinical psychology. These assessors were blind to treatment condition and had received a SCID training in accordance with their individual level of expertise. The assessor at preassessment was in most cases a different person than the therapist (52/60). In a minority of cases (8/60), the assessor became also the patient's therapist after the assessment. Note, however, that these assessors were also blind to condition because condition allocation took place after the preassessment. The number of administered SCID-I modules depended on participants' responses to the SCID-I screening questions (covering substance use disorders, anxiety disorders, and eating disorders). The modules on social phobia, mood disorders, psychotic disorders, post-traumatic stress disorder, and somatoform disorders were assessed for all patients. The avoidant personality disorder section of the Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II; First, Gibbon, Spitzer, Williams, & Benjamin, 1997) was also administered

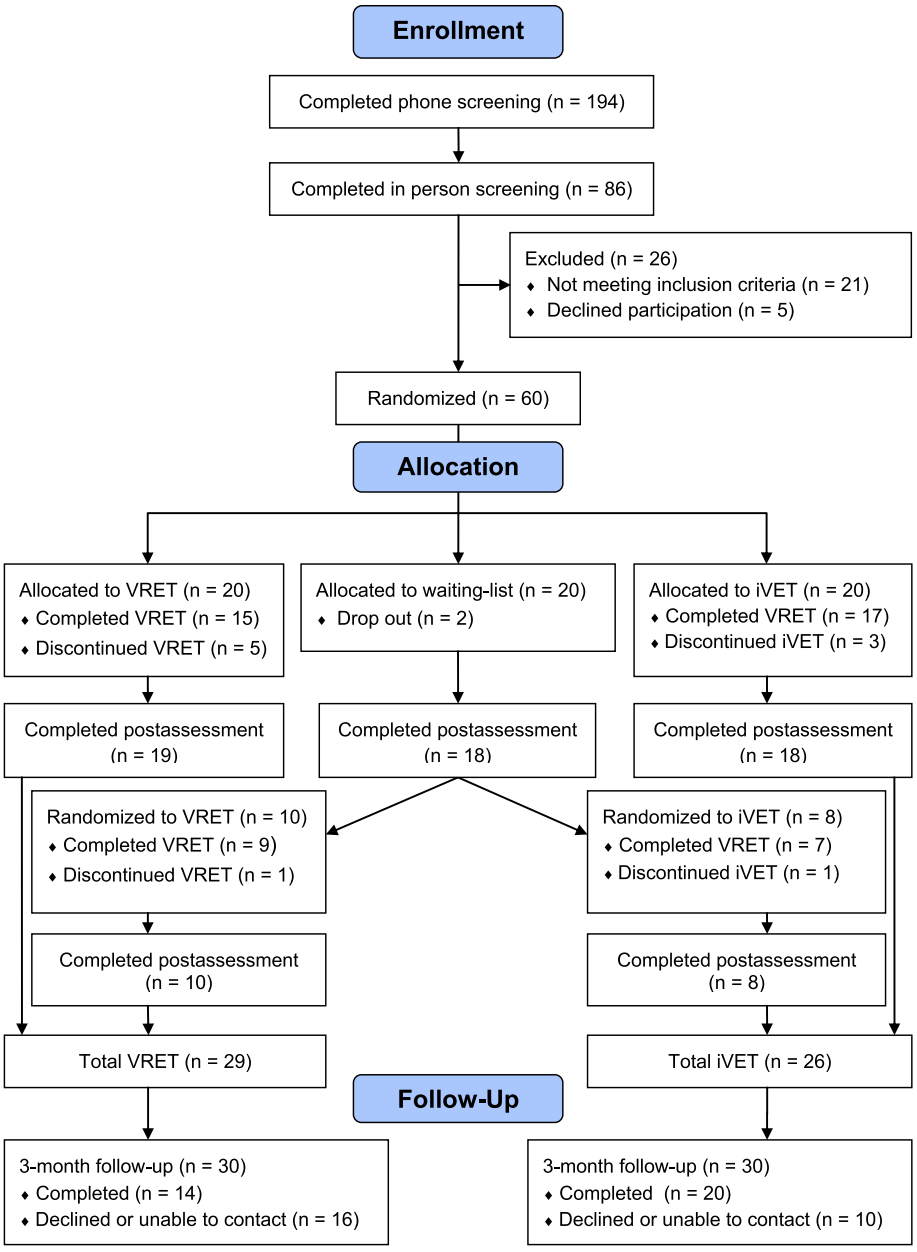

Fig. 1. Participant flow chart. VRET = Virtual Reality Exposure Therapy; iVET = in Vivo Exposure Therapy.

because research suggests that avoidant personality disorder and SAD might be one disorder instead of two distinct disorders, with avoidant personality disorder being the more severe form (Reich, 2009).

1.2.2. Primary outcome measures

Social anxiety symptoms were measured with the Liebowitz Social Anxiety Scale-Self Report (LSAS-SR; Liebowitz, 1987). The LSAS-SR is a 24-item questionnaire that assesses fear and avoidance in social situations on a 4-point Likert scale. The 12-week test-retest reliability of the LSAS-SR has been reported to be high ($r = 0.83$; Baker, Heinrichs, Kim, & Hofmann, 2002) and the internal consistency in the present study was excellent (Cronbach's $\alpha = .90–0.97$).

The subjective fear of being negatively evaluated by others in social situations was assessed with the Fear of Negative Evaluation Scale-Brief Form (FNE-B; Leary, 1983). The FNE-B is a 12-item instrument using a 5-point Likert scale for responses. Good psychometric properties have been reported for the FNE-B in earlier research (Weeks et al., 2005) and the internal consistency in the present study was excellent (Cronbach's $\alpha = .91–0.97$).

1.2.3. Secondary outcome measures

We measured speech duration and speech performance during a behavioural assessment task, in the form of a 5 min impromptu speech, to evaluate levels of behavioural avoidance. The behavioural assessment task was a modified version of a standardized protocol (Beidel, Turner, & Jacob, 1989). This modified version has been used in previous studies on social anxiety (Amir, Weber, Beard, Bomyea, & Taylor, 2008). Although participants with diverse social fears were included in the present study, this task was chosen because public speaking anxiety is the most prevalent

Table 1
Demographic characteristics of participants per condition.

Characteristics	VRET ($n = 20$)	iVET ($n = 20$)	WL ($n = 20$)
Age, *M (SD)*	39.65 (11.77)	37.50 (11.27)	33.50 (11.44)
Gender (% female)	65	75	50
Native language, *n (%)*			
Dutch	17 (85)	17 (85)	20 (100)
Spanish	1 (5)	0 (0)	0 (0)
Russian	1 (5)	0 (0)	0 (0)
Portuguese	0 (0)	1 (5)	0 (0)
Polish	0 (0)	1 (5)	0 (0)
Indonesian	0 (0)	1 (5)	0 (0)
Berber	1 (5)	0 (0)	0 (0)
Education, *n (%)*			
High	8 (40)	10 (50)	11 (55)
Middle	11 (55)	8 (40)	9 (45)
Low	1 (5)	2 (10)	0 (0)
Employment status, *n* (%)			
Paid employment	10 (50)	13 (65)	13 (65)
Trainee/student	1 (5)	1 (5)	5 (25)
Social welfare	1 (5)	1 (5)	0 (0)
Unemployed with voluntary work	1 (5)	0 (0)	0 (0)
Unemployed	7 (35)	5 (25)	2 (10)
Marital status, *n (%)*			
Married or cohabitating	9 (45)	10 (50)	11 (55)
Long distance relationship	2 (10)	3 (15)	2 (10)
Single living with children	1 (5)	0 (0)	0 (0)
Single living without children	7 (35)	7 (35)	6 (30)
Widowed	1 (5)	0 (0)	1 (5)
Comorbidity, *n* (%)			
Any anxiety disorder	3 (15)	4 (20)	0 (0)
Depressive disorder	4 (20)	0 (0)	2 (10)
Avoidant personality disorder	7 (35)	6 (30)	3 (15)
Session completed, *n*			
1	20	20	
2	20	20	
3	19	19	
4	19	19	
5	18	17	
6	15	17	
7	15	17	
8	15	16	
9	15	16	
10	14	14	
Dropout, *n* (%)	5 (25)	3 (15)	4 (20)

Note. VRET = Virtual Reality Exposure Therapy; iVET = in Vivo Exposure Therapy; WL = waiting-list; Low = completed elementary school or lower vocational education; Middle = completed high school or middle-level vocational education; High = completed pre-university, college, or university degree.

social fear. Speech duration was measured using a stop watch. To assess speech performance, two independent judges, blind for condition and assessment point, rated the videotaped speeches using 17 items of a public speaking performance measure on a 5-point Likert scale (Rapee & Lim, 1992). Higher scores on this measure indicated better speech performance. The internal consistency of this scale was good in earlier research ($r = 0.84$; Rapee & Lim, 1992) and the present study ($r = 0.81$–0.87).

Symptoms of depression, general anxiety, and stress were measured with the Depression Anxiety Stress Scale (DASS-21; Lovibond & Lovibond, 1995). The DASS-21 is a 21-item self-report questionnaire measuring depression, anxiety, and stress on a 4-point Likert scale with higher scores representing higher levels of depression, anxiety, or stress, respectively. The stress scale includes items that measure subjective coping with stressful events, such as tension, irritability, and a tendency to overreact to stressful events. The DASS possesses good psychometric properties (Antony, Bieling, Cox, Enns, & Swinson, 1998; Henry & Crawford, 2005) and in the present study its internal consistency was excellent (Cronbach's $\alpha = .91$–0.95).

Avoidant personality disorder related beliefs were assessed with the Personality Disorder Belief Questionnaire (PDBQ; Dreessen & Arntz, 1995). Research has shown that exposure therapy without cognitive components can affect cognitions (Powers et al., 2008). The avoidant subscale of the PDBQ contains 10 items to assess the strength of beliefs assumed to be specific to avoidant personality disorder. The internal consistency of this subscale was excellent in the present study (Cronbach's $\alpha = .90$–0.97).

Subjective quality of life was measured using the Eurohis Quality of Life Scale (EUROHIS-QOL 8-item index; Schmidt, Mühlan, & Power, 2006). The EUROHIS-QOL 8-item index measures quality of life on a 5-point Likert scale with higher scores indicating a better quality of life. The psychometric properties of the EUROHIS-QOL are reported to be satisfactory (Da Rocha, Power, Bushnell, & Fleck, 2012; Schmidt et al., 2006) and the internal consistency in the present study was good (Cronbach's $\alpha = .83$–0.93).

1.3. Procedure

The present study was approved by the Institutional Review Board of the University of Amsterdam and registered (NCT01746667; www.clinicaltrials.gov). Potential participants were asked on the telephone about former SAD treatment and whether attending treatment was logistically feasible. Afterwards, they filled in the SIAS online. Participants who scored above the cut-off were invited to an in-person intake (SCID), where they were screened for

exclusion criteria. After obtaining informed consent, eligible participants underwent a preassessment comprising a battery of self-report measures (LSAS-SR, FNE-B, DASS-21, PDBQ, EUROHIS-QOL) and the behavioural assessment task. For the behavioural assessment task, participants were told that they would give a 5 min speech in front of a camera and a two-person jury rating the speech. They were then asked to choose one out of seven topics (nuclear power, gay marriage, euthanasia, republic or monarchy, genetic selection, burqa ban, or mandatory organ donation) and had 2 min to prepare the speech. Participants were allowed to make notes during the preparation time but they could not use them during the speech. Then, the jury entered the room and the participants gave a speech for 5 min or until they indicated that they wanted to stop. After the assessment, participants were randomized to one of the three conditions (VRET, iVET, or waiting-list) using a computerized random number generator (http://www.randomization.com). A person who was not involved in the present study kept a list with the randomization sequence in a locked office cupboard and prepared sealed envelopes containing the condition allocation. The first author opened the envelopes after participants were enrolled. Participants in the waiting-list condition received a second assessment after a waiting period of five weeks (i.e. the same aimed length of time as the treatment) before being randomized to one of the treatment conditions. After the last treatment session, all participants completed a postassessment identical to the preassessment. Three months after the post-assessment, participants were invited to an in-person follow-up assessment consisting of the battery of self-report measures (LSAS-SR, FNE-B, DASS-21, PDBQ, EUROHIS-QOL).

1.4. Treatment

The treatment protocols for VRET and iVET were based on the protocols of Scholing and Emmelkamp (1993) and Hofmann and Otto (2008). Consistent with our aim of examining the potential efficacy of exposure to virtual social interactions, only behavioural exposure elements were used in both conditions and cognitive elements were discarded. Both treatments comprised ten 90 min sessions scheduled twice a week. In standard treatment, homework is commonly added to therapy sessions. However, due to the technical equipment necessary for VRET, virtual exposure could only be implemented in the lab. Therefore, homework assignments were not feasible in this condition. To keep the amount of exposure equal in both conditions, no homework assignment was given in either condition and therapists were instructed not to encourage participants to practice exposure outside of therapy sessions. Therapists involved in the present study were clinical psychologists and students in their last semester of a clinical master's degree program. They received training on VRET and iVET by the second and last author prior to administering both treatments. To monitor treatment adherence and competence, all therapy sessions and exposure exercises were extensively discussed during supervision. Furthermore, therapists were asked to complete a checklist immediately after each session in which they indicated any possible deviations from the protocol which were discussed during supervision. Moreover, therapy sessions were audio recorded (except for in vivo exposure exercises) and parts of recordings were replayed and discussed during supervision. Due to logistical reasons treatment adherence and competence were not formally assessed. Weekly supervision was provided to the therapists by the first, second, and last author.

1.4.1. Virtual reality exposure therapy (VRET)

VRET took place in the virtual reality laboratory of the University of Amsterdam. The laboratory consisted of two rooms separated by a one-way mirror, through which the therapist could see the participant during exposure exercises while controlling the computer system, whereas the participant could not see the therapist. The therapist and the participant had face-to-face contact before and after exposure exercises and during exposure they communicated via an intercom. The virtual situations covered one-to-one and group situations designed to provoke anxiety in individuals with SAD: giving a talk in front of an audience followed by questions from the audience, talking to a stranger, buying and returning clothes, attending a job interview, being interviewed by journalists, dining in a restaurant with a friend, and having a blind date (see Appendix A for a detailed description of all virtual scenarios and Figure 2 in Hartanto et al. (2014) for pictures of the virtual blind date, virtual job interview, and neutral world).

For virtual exposure, we used the Delft Remote Virtual Reality Exposure Therapy (DRVRET; Brinkman et al., 2012) system with virtual worlds which were visualized using a Vizard v3.0 software package. The setup consisted of three computers. The first computer, a custom Dell T3400 workstation, was used to run the VR server and the data logging system. The second computer, a custom Dell T3600 workstation using Intel Quadcore E5 with NVIDIA Quadro 5000, was used to run the VR engine and environment and the therapist could see simultaneously what the participant could see in the head mounted display. The video output of this computer was split for both the head mounted display (participant) and real time monitoring purpose (therapist). On the third computer, a custom Dell T3400 workstation, the therapist controlled the virtual situations. Participants wore a nVisor SX head mounted display with 1280 $\times$ 1024 pixels, a stereographic projection, and a 60° diagonal field of view.

Semi-structured dialogues controlled by the therapist ensured a certain length and difficulty level of interaction between the virtual humans and the participant, as well as allowing for individual responses for each participant (Brinkman et al., 2012). To tailor exposure exercises to the specific needs, anxiety level, and treatment progress of the individual participant, the system allowed the therapist to vary the following components depending on the virtual situation: dialogue style (friendly or unfriendly), gender of avatar, number of avatars present in the virtual world, dialogue topic's degree of personal relevance, and avatar's gestures (i.e., gaze direction and posture).

Treatment Sessions 1 and 2 focused on the conveyance of the therapy rationale, the registration of participant's relevant social situations, and creating a hierarchy of the available virtual social situations according to the participants' anticipated anxiety level. Moreover, participants were introduced to virtual reality and the technological equipment by entering a virtual neutral situation (Busscher, de Vliegher, Ling, & Brinkman, 2011), without any social interaction, for a maximum of 5 min.

Sessions 3 through 9 contained two 30 min blocks of exposure exercises separated by a 5 min break. The content of exposure exercises followed the previously made hierarchy in ascending order with regards to individual anxiety level (i.e. gradual exposure). Participants rated their anxiety level regarding three time points in every exposure exercise: beginning, highest level during the exercise, and end. Participants practiced every virtual world at least once and until anxiety decreased. Yet, only a maximum of two sessions were spent on exposure exercises focussing on presentation situations to limit the amount of practice in presentation performance, given that the behavioural assessment task also consisted of giving a speech. Session 10 was devoted to relapse prevention and evaluation of the therapy.

1.4.2. In vivo exposure therapy (iVET)

The iVET consisted of gradual exposure therapy to real-life

situations. Similar to VRET, iVET comprised 10 sessions with 60 min exposure in Sessions 3 through 9. As in the VRET condition, the therapy rationale and anxiety hierarchy were discussed in Sessions 1 and 2. The hierarchy used in iVET comprised participants' individual social situations which were translated to exposure exercises that could be implemented at the ambulatory of the University of Amsterdam or in its neighbourhood (e.g., supermarkets, subway stations, cafés, etc.). If relevant social situations could not be translated into exercises at the ambulatory or its nearby surroundings (e.g., work-related social situations), participants could substitute a regular session with a session in their personal environment. In these cases, the therapist and the participant had contact via the telephone before and after the exposure assignment. Session 10 was identical to the last session in the VRET condition.

1.5. Statistical analyses

Multilevel regression analyses were carried out to explore within-group (Time), between-group (Condition), and interaction (Time × Condition) effects. Only the fixed effects of the multilevel models were reported because they model change at the group level (in contrast to random effects, which model at the individual level). To investigate treatment effects from pre-to postassessment, each active treatment group was compared to waiting-list. The estimated model (see Table 3a) consisted of two parameters for each group: one parameter estimating the mean level of the outcome variable at preassessment and a second parameter estimating the rate of change from pre-to postassessment. For the active treatment groups, the second parameter described the change from pre-to postassessment relative to the change of the waiting-list group. To investigate long term effects of the two active treatments, the change from pre-to postassessment and the change from preassessment to follow-up were compared between VRET and iVET. The estimated model (see Table 3b) consisted of three parameters for each group: the first parameter describes the mean level of the outcome variable at preassessment, the second parameter describes the rate of change from pre-to postassessment, and the third parameter describes the rate of change from preassessment to follow-up. For VRET, all parameters described the estimations relative to iVET. Therefore, follow-up results only indicate whether the preassessment to 3-month follow-up change of VRET significantly differed from the iVRET change during the same period. Cohen's *d* estimations (Table 2) were based on the mean pre-to postassessment change of the active treatment group minus the mean pre-to postassessment change of the waiting-list group, divided by the pooled preassessment standard deviation (Morris, 2008). According to a power calculation (two-sided, power = 80%, alpha = 0.05; G*Power 3.1) based on an effect size of exposure therapy in a meta-analysis (Powers et al., 2008), 22 participants in each condition were sufficient to detect differences between treatment groups and waiting-list. Due to logistical reasons, data collection was discontinued after two years which resulted in two participants less in each condition. Participants completing less than six sessions were considered dropouts. In the VRET condition, one participant discontinued treatment due to motion-sickness. However, the two treatment groups did not significantly differ regarding number of participants who completed each session. There were no significant differences between VRET, iVET, and waiting-list regarding dropout rates (25%, 15%, and 20%, respectively). Treatment dropouts and completers did not significantly differ on demographic characteristics or outcome measures at preassessment (*ps* > 0.05). In line with the intent-to-treat paradigm, all participants who had started treatment (i.e., also participants who dropped out of treatment) were invited to complete the postassessment and follow-up assessment (Newell, 1992). Therefore, the number of participants who provided postassessment data was higher than the number of completers in both VRET and iVET (see Fig. 1). Analyses were carried out with both the intent-to-treat and the completer sample. Given that the completer sample did not significantly differ from the intent-to-treat sample on any of the variables, only the intent-to-treat data are reported in this article. For analyses comparing the two active treatment groups on the three time points, data of participants who received treatment directly and data of participants who received treatment after the waiting-list period, were pooled. The same analysis conducted with the unpooled data yielded a similar pattern of results. Analyses of the speech performance data were based on the average ratings of the two judges because inter-rater reliability was found to be high (Cronbach's $\alpha = .89$). The evaluation of assumptions showed that, for the FNE-B and speech duration data, the assumption of normally distributed error terms was violated. Square root transformation was applied in both cases but transformed- and untransformed data showed a similar pattern of results. Clinically significant improvement was defined as statistically reliable change on either of the primary outcome measures (LSAS-SR or FNE-B) and calculated according to Jacobson and Truax (1991) for participants who provided pre- and postassessment data.

2. Results

Analyses of variance, chi-squares and Fisher's exact tests revealed no significant differences between VRET, iVET, and waiting-list on demographic characteristics and outcome measures at preassessment (*ps* > 0.05) apart from perceived stress ($p = .011$). The results of the multilevel regression analyses are presented in Table 3. For descriptive statistics see Table 2.

2.1. Social anxiety

Multilevel regression analyses on LSAS-SR yielded a significant decrease from pre-to postassessment for both VRET ($p = .014$; $d = 0.55$) and iVET ($p < .001$; $d = 1.14$) compared to the waiting-list control group. However, the pre-to postassessment change significantly differed between VRET and iVET ($p = .006$) indicating a greater decrease for iVET than for VRET. The pre-to 3-month follow-up change of iVET was significant ($p < .001$). However, the change from preassessment to 3-month follow-up significantly differed between VRET and iVET ($p = .001$), indicating a greater decrease for iVET than for VRET (see Table 2).

Multilevel regression analyses on FNE-B revealed a significant decrease from pre-to postassessment for iVET ($p < .001$; $d = 1.60$) compared to the waiting-list control group. No significant differences were found between VRET and the waiting-list control group ($p = .197$). Furthermore, significant differences between VRET and iVET were observed for the pre-to postassessment changes ($p = .008$). The change from preassessment to 3-month follow-up was significant for iVET ($p < .001$). The pre-to 3-month follow-up change of VRET significantly differed from the iVET change ($p = .007$), indicating a greater decrease for iVET than for VRET.

2.2. Behavioural assessment task

Multilevel regression analyses on speech duration yielded a significant increase from pre-to postassessment for both VRET ($p = .018$; $d = 0.56$) and iVET ($p = .002$; $d = 0.77$) compared to the waiting-list control group. No significant differences were found between VRET and iVET at postassessment ($p = .920$).

Multilevel regression analyses on speech performance revealed

Table 2
Means and standard deviations for preassessment, postassessment, and 3-month follow-up with corresponding Cohen's ds.

Measures	Group	Pre		Post		3-month follow-up		
		M	*SD*	*M*	*SD*	*M*	*SD*	*d*
LSAS-SR	Virtual reality exposure	73.00	17.25	55.74	18.65	57.89	23.60	0.55
	In vivo exposure	69.15	19.44	39.22	25.01	29.92	22.22	1.14
	Waiting-list	64.35	21.87	58.00	16.01			
FNE-B	Virtual reality exposure	41.05	7.48	36.05	8.37	34.67	8.15	0.47
	In vivo exposure	38.50	6.24	24.72	12.10	20.00	12.34	1.60
	Waiting-list	37.40	9.42	36.44	8.77			
Speech duration (sec)	Virtual reality exposure	219.94	104.65	248.39	97.06			0.56
	In vivo exposure	222.21	98.78	269.88	68.36			0.77
	Waiting-list	252.84	98.97	224.33	106.83			
Speech performance	Virtual reality exposure	48.32	5.79	51.76	5.77			0.31
	In vivo exposure	46.47	4.58	52.78	5.40			0.76
	Waiting-list	51.26	8.47	52.44	7.52			
DASS depression	Virtual reality exposure	9.70	5.48	8.37	5.38	8.22	5.65	0.03
	In vivo exposure	6.15	4.90	3.17	3.42	3.54	3.86	0.35
	Waiting-list	7.30	5.46	6.11	4.70			
DASS anxiety	Virtual reality exposure	9.70	4.32	7.21	4.26	6.44	3.84	0.32
	In vivo exposure	7.00	5.24	3.61	3.47	3.85	3.58	0.47
	Waiting-list	6.30	5.09	5.33	4.09			
DASS stress	Virtual reality exposure	12.55	5.10	9.47	4.30	10.89	5.71	0.44
	In vivo exposure	8.55	4.39	5.22	3.44	4.85	3.24	0.52
	Waiting-list	8.35	4.82	7.44	4.10			
EUROHIS-QOL	Virtual reality exposure	24.80	6.07	25.74	6.85	25.00	9.12	0.15
	In vivo exposure	26.90	4.51	29.78	4.47	28.92	6.46	0.55
	Waiting-list	27.20	5.67	27.28	5.44			
PDBQ	Virtual reality exposure	62.18	19.74	48.62	20.76	53.11	26.54	0.74
	In vivo exposure	47.47	17.07	28.08	20.19	22.90	17.61	1.05
	Waiting-list	49.73	24.64	52.64	22.03			

Note. LSAS-SR = Liebowitz Social Anxiety Scale-Self Report; FNE-B = Fear of Negative Evaluation Scale-Brief Form; DASS = Depression Anxiety Stress Scale; EUROHIS-QOL = Eurohis Quality of Life Scale; PDBQ = Personality Disorder Belief Questionnaire; Pre = preassessment; Post = postassessment.

a significant increase from pre-to postassessment for iVET ($p = .003$; $d = 0.76$) compared to the waiting-list control group. No significant differences for the pre-to postassessment change were found between VRET and the waiting-list control group ($p = .134$) or VRET and iVET ($p = .091$). There was no significant change from pre-to postassessment observed for the waiting-list control group.

2.3. Avoidant personality disorder

Multilevel regression analyses on avoidant personality disorder related beliefs, as measured with the PDBQ, revealed a significant decrease from pre-to postassessment for both VRET ($p = .002$; $d = 0.74$) and iVET ($p < .001$; $d = 1.05$) compared to the waiting-list control group. The pre-to 3-month follow-up change of iVET was significant ($p < .001$). While there was no significant difference observed in change from pre-to postassessment ($p = .134$) between VRET and iVET, there was a significant difference between these two groups in change from preassessment to 3-month follow-up ($p = .003$) indicating a greater decrease for iVET than VRET. There was no significant change from pre-to postassessment observed for the waiting-list control group.

2.4. Depression, general anxiety, perceived stress, and quality of life

Multilevel regression analyses on the DASS-21 stress subscale yielded a significant decrease from pre-to postassessment for both VRET ($p = .032$; $d = 0.44$) and iVET ($p = .022$; $d = 0.52$) compared to the waiting-list control group. A significant decrease on the anxiety subscale ($p = .043$; $d = 0.47$) was observed for iVET compared to the waiting-list control group. No significant differences were found between VRET and waiting-list on the anxiety subscale ($p = .135$) and depression subscale ($p = .401$) or between iVET and waiting-list on the depression subscale ($p = .103$). The pre-to 3-month follow-up change of iVET was significant for the depression ($p = .008$), anxiety ($p < .001$), and stress subscales ($p < .001$). The decrease from pre-to postassessment and from preassessment to 3-month follow-up did not differ significantly between VRET and iVET for any of the three subscales ($p > .05$). There was no significant change observed for the waiting-list control group on any of the three subscales from pre-to postassessment.

Multilevel regression analyses on the EUROHIS-QOL yielded a significant increase from pre-to postassessment for iVET ($p = .001$; $d = 0.55$) compared to the waiting-list control group. No significant differences were found between VRET and the waiting-list control group ($p = .279$). The iVET change from pre-to 3-month follow-up was significant ($p = .010$). While VRET and iVET significantly differed in increase from pre-to postassessment ($p = .004$), they did not differ in change from preassessment to 3-month follow-up ($p = .253$). There was no significant pre-to postassessment change observed on this measure for the waiting-list control group.

2.5. Clinical improvement

Reliable change from pre-to postassessment was observed in 47.4% ($n = 9/19$) of participants who had received VRET, 77.8% ($n = 14/18$) who had received iVET, and 50.0% ($n = 9/18$) in the waiting-list condition, χ^2 (2, $N = 55$) = 4.25, $p = .119$. There was no significant difference between iVET and waiting-list ($p = .083$) or VRET and waiting-list ($p = .873$). The difference between VRET and iVET did not reach statistical significance ($p = .057$).

3. Discussion

The present study investigated the efficacy of VRET as a stand-alone treatment for individuals with SAD comprising heterogeneous social fears. The results revealed that participants in both VRET and iVET improved from pre-to postassessment on social anxiety, avoidance, speech duration during a behavioural

Table 3
Multilevel regression analyses for a) VRET, iVET, and waiting-list on all measures at pre- and postassessment and b) VRET and iVET on all measures at preassessment, postassessment, and 3-month follow-up.

	LSAS-SR		FNE-B		Speech duration (sec)		Speech performance		DASS depression		DASS anxiety		DASS stress		EUROHIS-QOL		PDBQ	
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
a)																		
VRET baseline	73.00	4.27^{***}	41.05	1.71^{***}	222.71	22.68^{***}	48.32	1.44^{***}	9.70	1.15^{***}	9.70	1.07^{***}	12.55	1.04^{***}	24.80	1.19^{***}	62.18	4.52^{***}
iVET baseline	69.15	4.27^{***}	38.50	1.71^{***}	222.21	22.36^{***}	46.29	1.46^{***}	6.15	1.15^{***}	7.00	1.07^{***}	8.55	1.04^{***}	26.90	1.19^{***}	47.47	4.52^{***}
WL baseline	64.35	4.27^{***}	73.40	1.71^{***}	251.59	22.09^{***}	50.79	1.48^{***}	7.30	1.15^{***}	6.30	1.07^{***}	8.35	1.04^{***}	27.20	1.19^{***}	49.73	4.52^{***}
WL pre-post Δ	−3.88	3.90^{ns}	−0.98	2.19^{ns}	−33.58	17.63^{ns}	0.94	1.17^{ns}	−0.73	0.99^{ns}	−0.67	0.95^{ns}	−0.37	0.93^{ns}	0.06	0.61^{ns}	3.29	3.68^{ns}
WL pre-post Δ × VRET	−13.92	5.46^{*}	−3.99	3.06^{ns}	61.18	25.01^{*}	2.42	1.59^{ns}	−0.76	1.39^{ns}	−1.69	1.33^{ns}	−2.85	1.30^{*}	0.94	0.86^{ns}	−16.49	5.14^{**}
WL pre-post Δ × iVET	−27.19	5.52^{***}	−12.56	3.09^{***}	82.68	25.2^{**}	5.07	1.63^{**}	−2.32	1.40^{ns}	−2.77	1.34^{*}	−3.10	1.31^{*}	3.08	0.87^{**}	−22.11	5.21^{***}
b)																		
Intercept	65.64	3.45^{***}	38.00	1.42^{***}	241.37	18.99^{***}	48.67	1.23^{***}	6.32	0.97^{***}	6.50	0.90^{***}	8.18	0.90^{***}	26.82	1.01^{***}	48.39	3.65^{***}
VRET	2.66	4.80^{ns}	1.43	1.98^{ns}	−35.68	26.51^{ns}	0.18	1.72^{ns}	2.01	1.34^{ns}	1.77	1.25^{ns}	2.72	1.25^{*}	−1.02	1.40^{ns}	11.12	5.07^{*}
iVET pre-post Δ	−29.94	3.38^{***}	−12.00	1.87^{***}	40.05	15.14^{*}	5.11	0.95^{***}	−2.53	0.80^{**}	−3.13	0.77^{***}	−2.78	0.83^{**}	2.80	0.49^{***}	−19.14	3.08^{***}
iVET pre-post Δ × VRET	13.42	4.66^{**}	7.03	2.57^{**}	2.11	20.80^{ns}	−2.27	1.31^{ns}	1.67	1.11^{ns}	1.66	1.06^{ns}	0.55	1.15^{ns}	−2.01	0.68^{**}	6.45	4.24^{ns}
iVET pre-fu Δ	−35.01	3.82^{***}	−14.82	2.13^{***}					−2.41	0.85^{**}	−3.17	0.73^{***}	−3.32	0.75^{***}	2.15	0.79^{*}	−24.42	3.30^{***}
iVET pre-fu Δ × VRET	20.06	5.65^{**}	8.89	3.16^{**}					1.22	1.29^{ns}	1.52	1.07^{ns}	1.41	1.12^{ns}	−1.37	1.18^{ns}	15.37	4.95^{**}

Note. LSAS-SR = Liebowitz Social Anxiety Scale-Self Report; FNE-B = Fear of Negative Evaluation Scale-Brief Form; DASS = Depression Anxiety Stress Scale; EUROHIS = Eurohis Quality of Life Scale; PDBQ = Personality Disorder Belief Questionnaire; Pre = preassessment; Post = postassessment; Fu = 3-month follow-up; VRET = Virtual Reality Exposure Therapy; iVET = in Vivo Exposure Therapy; WL = waiting-list.
$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

assessment task, perceived stress, and avoidant personality disorder related beliefs when compared to the waiting-list control group. Participants receiving iVET, but not VRET, improved on fear of negative evaluation, speech performance, general anxiety, depression, and quality of life relative to waiting-list. Additionally, iVET was superior to VRET regarding decrease on both social anxiety symptoms at post- and follow-up assessment and avoidant personality disorder related beliefs at follow-up. Speech performance only improved for iVET when compared to waiting-list, but iVET did not differ from VRET at postassessment. For VRET, only effects on perceived stress were significant from preassessment to 3-month follow-up. For iVET, all improvements were also significant at 3-month follow-up. Clinically significant improvement regarding social anxiety and avoidance, or fear of negative evaluation did not significantly differ between the VRET, iVET, and waiting-list.

These results are in line with earlier randomized controlled trials (i.e., Anderson et al., 2013; Bouchard et al., 2015; Wallach et al., 2009) that found VRET to be effective in treating social anxiety complaints. However, our results extend these findings by showing that (1) VRET is effective without the addition of cognitive components, (2) for participants with heterogeneous social fears, (3) comparing both VRET and iVET in individual format, (4) as well as including extensive virtual verbal interaction. Accordingly, our findings suggest that VRET comprising diverse virtual situations and verbal interaction is potentially effective for individuals with generalized SAD. Moreover, the two treatments did not differ regarding their effect on speech duration during the behavioural assessment task. This suggests that VRET might be comparable to iVET when looking at actual behavior. The fact that speech duration did not change in the waiting-list control group suggests that the improvement of the active treatment groups cannot be attributed to repeated exposure to the behavioural assessment task. However, the results on speech performance as rated by independent judges were less conclusive. Although there was no difference between the two active conditions, only iVET proved to be effective in improving speech performance when compared to the waiting-list control group. Still, these results indicate that VRET has the potential to produce effects that can generalize to real-life social situations. Additionally, our results indicate that VRET effectively reduces beliefs related to avoidant personality disorder. These beliefs are not merely situation specific; they contain general cognitions about the social-self and others. However, this effect was significant at post-treatment only. In line with Anderson et al. (2013) and Bouchard et al. (2015), dropout rates did not significantly differ between VRET, iVET, and waiting-list. In summary, the present trial indicates that VRET as a stand-alone treatment is effective, and that virtual verbal interaction can successfully be applied for treatment purposes in individuals with SAD.

Nevertheless, our virtual reality intervention was less effective than iVET with regards to several measured domains and results were less conclusive than those of previous studies (Anderson et al., 2013; Bouchard et al., 2015; Wallach et al., 2009). We believe that this finding needs to be interpreted in the context of our study representing the first attempt to develop and apply diverse and complex virtual social interactions to target diverse social fears. Furthermore, VRET was administered as a standalone treatment without any additional therapeutic component. Although VRET effectively reduced anxiety and avoidance in social situations, it did not significantly reduce fear of negative evaluation, which represents a cognitive core feature of SAD. Pure exposure therapy can lead to cognitive changes in SAD (Powers et al., 2008) and iVET significantly reduced fear of negative evaluation in the present study. Yet, the need to directly address cognitions might be higher during VRET than during iVET. Namely, during VRET, participants

can easily make use of cognitive avoidance (e.g., 'the virtual social world is not real so I do not need to be afraid'). However, our results are in line with the study by Anderson et al. (2013), in which fear of negative evaluation was also not significantly affected by VRET. Furthermore, in their study, VRET was administered in combination with cognitive elements and still did not significantly reduce fear of negative evaluation. This suggests that the absence of cognitive elements in our study does not entirely explain why VRET did not significantly affect fear of negative evaluation. However, although CBT might not be more effective than pure exposure when administered in vivo (Powers et al., 2008), the addition of cognitive elements might improve the efficacy of VRET as indicated by the greater effect size in Anderson et al. (2013) compared to the present study.

Our study can be used to inform future research on applying virtual social interactions in treating SAD and other psychological complaints related to difficulties with social interactions. There are several recommendations that might improve VRET for SAD. First, in our trial, the dialogues were semi-structured to allow therapists to control the level of difficulty and content of the conversation between participants and virtual humans. Therefore, in contrast to in vivo exposure, the flexibility of the dialogues was limited within VRET. Considering research showing that individuals with SAD fear uncertainty (Boelen & Reijntjes, 2009; Carleton, Collimore, & Asmundson, 2010), more extensive and flexible dialogues might further improve the usability of VRET by allowing for more individualized responses, increased uncertainty, and therefore more realistic and unpredictable social interaction. Second, even though our virtual worlds addressed social situations, which the majority of individuals with SAD fear (Ruscio et al., 2008), the total number of virtual situations within the present study was rather limited. Furthermore, the stimuli that trigger social anxiety within a certain social situation might differ for each individual. Therefore, increasing the number of available virtual situations and creating the possibility of adjusting virtual worlds to the individual's needs by adding or removing social cues would allow individuals to practice more personally relevant situations. A third recommendation is related to facial expressions of virtual humans that were part of VRET. In our trial, fear of negative evaluation was not significantly affected by VRET. The limited effect ($d = 0.45$) might be explained by the absence of facial expressions in the virtual humans. Facial expressions can reveal what others feel or think, and therefore might also be essential for the experience of being negatively evaluated. Anderson et al. (2013) included an audience displaying facial expressions in their study and still did not find an effect of VRET on fear of negative evaluation. However, facial expressions might play a more important role in one-to-one social interaction than when standing in front of an audience where facial expressions are diverse and the individual facial expression might be more difficult to register. Hence, adding facial expressions to virtual humans in VRET containing one-to-one social interaction might help trigger the fear of negative evaluation and increase sense of presence within VRET (Qu, Brinkman, Ling, Wiggers, & Heynderickx, 2014).

Although it is too early to conclude whether VRET can effectively be administered as a standalone treatment for SAD in clinical practice, the present study helped to make a step in this direction by showing that we are able to simulate intensive social interaction in virtual reality and that exposure to them affects social anxiety complaints. Incorporation of more extensive and flexible dialogues, a greater number of virtual scenarios, facial expressions, and cognitive elements into VRET may further improve treatment outcome. Moreover, several advantages might make VRET a valuable addition to existing treatments. First, for individuals who are not willing to participate in exposure in vivo because of their fear, VRET could represent a first step in the exposure hierarchy. With regard to specific phobias, research indicates that participants might prefer VRET to exposure in vivo. For example, treatment refusal rates for the former were lower compared with the latter (Garcia-Palacios, Botella, Hoffman, & Fabregat, 2007). Future research needs to investigate whether this might also apply to VRET for SAD. Second, during in vivo exposure, therapists often encounter difficulties associated with the planning and implementation of exposure exercises (e.g., giving a speech in front of an audience, talking to strangers) due to the unpredictable nature and possibly short duration of naturally occurring social interaction, as well as the need for human resources to realize the exercises. By simulating social situations, VRET allows therapists to control the content, duration, and difficulty of social interactions (Hartanto et al., 2014). Third, participant and therapist do not need to leave the therapist's office, which is associated with less treatment costs and assures participants' privacy throughout exposure exercises (Emmelkamp, 2005).

If technological and psychological improvements of VRET for SAD can enhance its treatment efficacy, future research should further focus on automated treatment. In the present study, the therapist controlled the system and the avatar's responses. Therefore, treatment could only take place within a clinical setting, where participant and therapist were at the same physical location. Developing software that uses speech recognition to identify responses and to select appropriate answers for the virtual human would automate the treatment process to a certain extent. This would allow treatment of patients who are too afraid to leave their house or who live in an area with no psychological treatment available, and could reduce treatment costs in the future. It could also make virtual reality homework assessments more feasible. The costs for virtual reality hardware have decreased considerably over the past years and new developments (e.g., virtual reality generating smartphones) might increase the access of VRET for the general population.

A limitation of the present study is that social situations in exposure exercises were not identical in both treatment conditions. Accordingly, participants in iVET could practice in a greater number and variety of social situations compared to VRET, which might explain the increased efficacy of iVET relative to VRET. In a future study, iVET and VRET with identical exposure exercises should be compared to rule out the number of social situations as an explanation for the difference in efficacy between iVET and VRET. A second limitation is the absence of homework exercises in both treatment conditions. In the present study, exposure at home was not feasible in the VRET condition due to logistical reasons and therefore, we did not include homework exercises in both conditions. However, homework exercises are an integral part of exposure therapy in clinical practice and both treatment conditions might profit from its inclusion in treatment.

In conclusion, VRET as a stand-alone treatment with virtual verbal interaction can produce significant therapeutic gains in participants with SAD. Future research needs to focus on improving technological and psychological aspects of virtual social interactions to improve the overall treatment efficacy of virtual reality based interventions.

Acknowledgements

This project (055.10.207) was funded by the Netherlands Organisation for Scientific Research (NWO). The authors have no financial interests or potential conflicts of interest.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.brat.2015.12.016.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Amir, N., Weber, G., Beard, C., Bomyea, J., & Taylor, C. T. (2008). The effect of a single-session attention modification program on response to a public-speaking challenge in socially anxious individuals. *Journal of Abnormal Psychology, 117*(4), 860–868.

Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. A., Schmertz, S. K., Zimand, E., et al. (2013). Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial. *Journal of Consulting and Clinical Psychology, 81*(5), 751–760.

Anderson, P., Rothbaum, B. O., & Hodges, L. F. (2003). Virtual reality exposure in the treatment of social anxiety. *Cognitive and Behavioral Practices, 10*, 240–247.

Anderson, P. L., Zimand, E., Hodges, L. F., & Rothbaum, B. O. (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety, 22*(3), 156–158.

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the depression anxiety stress scales in clinical groups and a community sample. *Psychological Assessment, 10*(2), 176–181.

Baker, S. L., Heinrichs, N., Kim, H. J., & Hofmann, S. G. (2002). The Liebowitz social anxiety scale as a self-report instrument: a preliminary psychometric analysis. *Behaviour Research and Therapy, 40*(6), 701–715.

Beidel, D. C., Turner, S. M., & Jacob, R. G. (1989). Assessment of social phobia: reliability of an impromptu speech task. *Journal of Anxiety Disorders, 3*, 149–158.

Boelen, P. . a., & Reijntjes, A. (2009). Intolerance of uncertainty and social anxiety. *Journal of Anxiety Disorders, 23*(1), 130–135.

Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, E., Forget, H., et al. (2015). *Using virtual reality to conduct exposure to a broad range of stimuli in the treatment of social anxiety disorder: A three-arm randomized control trial* (Manuscript Submitted for Publication).

Brinkman, W.-P., Hartanto, D., Kang, N., de Vliegher, D., Kampmann, I. L., Morina, N., et al. (2012). A virtual reality dialogue system for the treatment of social phobia. In *Proceedings of the 2012 ACM Annual Conference extended abstracts on human factors in computing systems extended abstracts – CHI EA'12, 1099*.

Busscher, B., de Vliegher, D., Ling, Y., & Brinkman, W. P. (2011). Physiological measures and self-report to evaluate neutral virtual reality worlds. *Journal of Cyber Therapy and Rehabilitation, 4*(1), 15–25.

Carleton, R. N., Collimore, K. C., & Asmundson, G. J. G. (2010). "It's not just the judgements-It's that I don't know": intolerance of uncertainty as a predictor of social anxiety. *Journal of Anxiety Disorders, 24*(2), 189–195.

Da Rocha, N. S., Power, M. J., Bushnell, D. M., & Fleck, M. P. (2012). The EUROHIS-QOL 8-item index: comparative psychometric properties to its parent WHOQOL-BREF. *Value in Health, 15*(3), 449–457.

Dreessen, L., & Arntz, A. (1995). *The personality disorder beliefs questionnaire (short version)*. Maastricht: Author.

Emmelkamp, P. M. G. (2005). Technological innovations in clinical assessment and psychotherapy. *Psychotherapy and Psychosomatics, 74*, 336–343.

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *Structured clinical interview for DSM-IV Axis II personality disorders*. Washington, D.C.: American Psychiatric Press, Inc.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1994). *Structured clinical interview for Axis I DSM–IV Disorders—Research ed*. New York, NY: Biometrics Research Department.

Garcia-Palacios, A., Botella, C., Hoffman, H., & Fabregat, S. (2007). Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society, 10*(5), 722–724.

Hartanto, D., Kampmann, I. L., Morina, N., Emmelkamp, P. M. G., Neerincx, M. A., & Brinkman, W. P. (2014). Controlling social stress in virtual reality environments. *PLoS One, 9*(3).

Heimberg, R. G., Mueller, G. P., Holt, C. S., Hope, D. A., & Liebowitz, M. R. (1992). Assessment of anxiety in social interaction and being observed by others: the social interaction anxiety scale and the social phobia scale. *Behavior Therapy, 23*(1), 53–73.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the depression anxiety stress scales (DASS-21): construct validity and normative data in a large non-clinical sample. *The British Journal of Clinical Psychology/the British Psychological Society, 44*(Pt 2), 227–239.

Hofmann, S. G., & Otto, M. W. (2008). *Cognitive-behavior therapy of social anxiety disorder: Evidence-based and disorder specific treatment techniques*. New York, NY: Routledge.

Hofmann, S. G., & Smits, J. A. J. (2008). Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry, 69*, 621–632.

Klinger, E., Bouchard, S., Legeron, P., Roy, S., Laurer, F., Chemin, I., et al. (2005). Virtual reality therapy versus cognitive behavior therapy for social phobia: a preliminary controlled study. *CyberPsychology & Behavior, 8*(1), 76–88.

Leary, M. R. (1983). A brief version of the fear of negative evaluation scale. *Personality and Social Psychology Bulletin, 9*(3), 371–375.

Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry, 22*, 141–173.

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales*. Sydney: Psychological Foundation Monograph.

Mattick, R., & Clarke, C. (1998). Development and validation of measure of social phobia scrutiny fear and social interaction anxiety. *Behavior Research and Therapy, 36*, 455–470.

Mayo-Wilson, E., Dias, S., Mavranezouli, I., Kew, K., Clark, D. M., Ades, A. E., et al. (2014). Psychological and pharmacological interventions for social anxiety disorder in adults: a systematic review and network meta-analysis. *The Lancet Psychiatry, 1*, 368–376.

Meyerbröker, K., & Emmelkamp, P. M. G. (2010). Virtual reality exposure therapy in anxiety disorders: a systematic review of process-and-outcome studies. *Depression and Anxiety, 27*(10), 933–944.

Morina, N., Brinkman, W.-P., Hartanto, D., & Emmelkamp, P. M. G. (2014). Sense of presence and anxiety during virtual social interactions between a human and virtual humans. *PeerJ, 2*, e337.

Morina, N., Ijntema, H., Meyerbröker, K., & Emmelkamp, P. M. G. (2015). Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. *Behaviour Research and Therapy, 74*, 18–24.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*, 364–386.

Newell, D. J. (1992). Intention-to-treat analysis: implications for quantitative and qualitative research. *International Journal of Epidemiology, 21*(5), 837–841.

Opriş, D., Pintea, S., García-Palacios, A., Botella, C., Szamosközi, Ş., & David, D. (2012). Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. *Depression and Anxiety, 29*(2), 85–93.

Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 39*(3), 250–261.

Powers, M. B., Briceno, N. F., Gresham, R., Jouriles, E. N., Emmelkamp, P. M. G., & Smits, J. A. J. (2013). Do conversations with virtual avatars increase feelings of social anxiety? *Journal of Anxiety Disorders, 27*(4), 398–403.

Powers, M. B., Sigmarsson, S. R., & Emmelkamp, P. M. G. (2008). A meta–analytic review of psychological treatments for social anxiety disorder. *International Journal of Cognitive Therapy, 1*(2), 94–113.

Qu, C., Brinkman, W. P., Ling, Y., Wiggers, P., & Heynderickx, I. (2014). Conversations with a virtual human: synthetic emotions and human responses. *Computers in Human Behavior, 34*, 58–68.

Rapee, R. M., & Lim, L. (1992). Discrepancy between self- and observer ratings of performance in social phobics. *Journal of Abnormal Psychology, 101*, 728–731.

Reich, J. (2009). Avoidant personality disorder and its relationship to social phobia. *Current Psychiatry Reports, 11*(1), 89–93.

Ruscio, A. M., Brown, T. A., Chiu, W. T., Sareen, J., Stein, M. B., & Kessler, R. C. (2008). Social fears and social phobia in the USA: results from the national comorbidity survey replication. *Psychological Medicine, 38*(1), 15–28.

Schmidt, S., Mühlan, H., & Power, M. (2006). The EUROHIS-QOL 8-item index: psychometric results of a cross-cultural field study. *European Journal of Public Health, 16*(4), 420–428.

Scholing, A., & Emmelkamp, P. M. G. (1993). Exposure with and without cognitive therapy for generalized social phobia: effects of individual and group treatment. *Behaviour Research and Therapy, 31*, 667–681.

Wallach, H. S., Safir, M. P., & Bar-Zvi, M. (2009). Virtual reality cognitive behavior therapy for public speaking anxiety: a randomized clinical trial. *Behavior Modification, 33*(3), 314–338.

Weeks, J. W., Heimberg, R. G., Fresco, D. M., Hart, T. A., Turk, C. L., Schneier, F. R., et al. (2005). Empirical validation and psychometric evaluation of the brief fear of negative evaluation scale in patients with social anxiety disorder. *Psychological Assessment, 17*(2), 179–190.

Wittchen, H.-U., Fuetsch, M., Sonntag, H., Müller, N., & Liebowitz, M. (2000). Disability and quality of life in pure and comorbid social phobia—findings from a controlled study. *European Psychiatry, 15*, 46–58.