Institutionen för klinisk neurovetenskap
Psykologprogrammet, termin 9-10
Huvudämne: Psykologi
Examensarbete i psykologi (2PS026), 30 högskolepoäng
Vårterminen 2016

Predictors of treatment outcome in exposure therapy for spider phobia

Maria Garke & Jonas Rafi

Handledare: Professor Per Carlbring, Psykologiska Institutionen vid Stockholms Universitet
Bihandledare: Professor Gerhard Andersson, Institutionen för klinisk neurovetenskap
Examinator: Professor Bo Melin, Institutionen för klinisk neurovetenskap

Institutionen för klinisk neurovetenskap
Psykologprogrammet, termin 9-10
Huvudämne: Psykologi
Examensarbete i psykologi (2PS026), 30 högskolepoäng
Vårterminen 2016

Predictors of treatment outcome in exposure therapy for spider phobia

Sammanfattning. Specifik fobi är den näst vanligaste psykiatriska diagnosen, och kan kraftigt påverka livskvaliteten negativt för den drabbade. Den rekommenderade behandlingen vid specifik fobi är exponeringsterapi. Under senare tid har användandet av virtuella verkligheter, eller Virtual Reality (VR), visat lovande resultat inom forskningen och kan vara ett potentiellt alternativ till traditionell exponeringsterapi. Studier har visat att exponeringsterapi med hjälp av VR kan vara effektivt, men mer forskning behövs gällande vad som kan predicera behandlingsutfall. Syftet med denna studie var att undersöka prediktorer för behandlingsutfall i en randomiserad, kontrollerad studie av en session med VR-behandling för spindelfobi, jämfört mot en aktiv kontrollbehandling (en-sessions exponeringsterapi). Resultaten visade att låg grad av fobiska symptom innan behandling predicerade bättre behandlingsresultat på det primära utfallsmåttet ($p < .001$). Förtroende för behandling predicerade bättre behandlingsresultat endast i VR-gruppen ($p < .01$). Denna studie visar att grad av fobiska symptom innan behandling är viktigt för behandlingsutfallet vid exponeringsterapi, där individer med lägre symptombörda får bättre resultat. Detta innebär att bedömandet av grad av fobiska besvär är viktigt, särskilt vid VR-exponeringsbehandling, för att kunna individanpassa typ av behandling.

Nyckelord: exponeringsterapi, prediktorer, specifik fobi, spindelfobi, Virtual Reality.

Abstract. Specific phobia is the second most common mental health disorder, and may largely impact the quality of life negatively for afflicted individuals. Exposure therapy has been established as the treatment of choice regarding specific phobia. The application of exposure therapy in a Virtual Reality (VR) setting has gained momentum scientifically as an alternative treatment. Studies have shown that VR exposure treatment is effective, but very little research has been done on predictors of treatment outcome. The aim of this study was to investigate predictors of treatment outcome in a randomized controlled study of VR one-session exposure therapy for spider phobia, against an active control treatment (one-session in vivo exposure). The results showed that initial impairment of phobia had a significant impact on the primary outcome measure in both groups ($p < .001$). Treatment credibility had a significant impact only in the VR treatment group ($p < .01$). This study shows that initial impairment of phobia is of importance in exposure therapy. This implies that individuals with less symptoms could benefit more from VR treatment, and that assessing initial impairment of phobia for individualizing treatment is important.

Keywords: exposure therapy, predictors, specific phobia, spider phobia, Virtual Reality.

Predictors of treatment outcome in exposure therapy for spider phobia

Maria Garke & Jonas Rafi

Introduction

Specific phobia is the second most common mental health disorder after major depression disorder (Kessler, Petukhova, Sampson, Zaslavsky, & Wittchen, 2012), and may severely impair the quality of life for the person suffering from this condition (Magee, Eaton, Wittchen, McGonagle, & Kessler, 1996).

Specific phobia is characterized by a pronounced fear or anxiety towards a specific object or a specific situation (e.g. heights, animals, syringes or blood). Criteria for specific phobia according to the Diagnostic and Statistical Manual for Mental Disorders 5th edition (APA, 2013) include that the phobic stimulus immediately evokes anxiety or fear, is actively avoided, or endured with great anxiety or fear. The anxiety or fear evoked when confronted with the phobic stimulus is not proportional to the sociocultural context or the actual danger the object or situation entails. Furthermore, the symptoms of specific phobia have to be sustained for a longer period of time and cause clinically significant impairment in everyday life (APA, 2013).

Among anxiety disorders, specific phobia is the most common mental health disorder with a lifetime prevalence of 12.5% (Kessler et al., 2005). Prevalence tends to be lower among older populations compared with younger populations (APA, 2013). Specific phobia is often comorbid with additional psychopathology. Individuals who develop specific phobia at a young age are at risk for developing other mental disorders. These include different anxiety disorders, depressive and bipolar disorders and personality disorders (Stinson et al., 2007).

Current consensus regarding the etiology of specific phobia states that there does not seem to be any single process or event causing specific phobia. Instead, specific phobia is a result of multiple factors contributing to the development of the disorder, including genetic influences, distorted cognitions, differences in temperament, experiential avoidance and negative learning experiences (Ollendick & Muris, 2015). Most specific phobias develop before the age of 10, although situational-specific phobias tend to emerge later in life. During childhood, the severity of specific phobia symptoms tends to fluctuate and spontaneous recovery can occur. If phobia obtained in childhood is sustained until adulthood, the symptoms tend to stabilize or even worsen over time if untreated (APA, 2013).

The different types of specific phobias are divided into five distinct subtypes: animals (e.g. snakes, spiders), natural environments (e.g. heights, water), situational (e.g. elevators, airplanes), blood-injection-injury and other types (e.g. loud sounds or costumed characters) (APA, 2013). Among the different varieties of specific phobias, animal type phobias are highly prevalent with an estimated 7.9 % proportion of the population suffering from this condition at a specific point in time (Fredrikson, Annas, Fischer, & Wik, 1996). Spider phobia is the most prevalent and most studied animal phobia among specific phobias. It is associated with substantial impairment due to the consequences of anxiety and avoidance in everyday life (Choy, Fyer, & Lipsitz, 2007).

A cognitive behavioral model aiming to describe the phenomena that is specific phobia has been proposed by Öst (2013). This model provides a theoretical explanation to the development and maintenance of phobic fear and symptoms. It is presented in Figure 1. The

model is based on a) avoidance of the phobic object and b) catastrophic beliefs as the factors that contribute to the continuation of symptoms and fear. The patient suffering from specific phobia will have a set of catastrophic beliefs about what is going to happen if they encounter the phobic object. Because of these beliefs, the patient will avoid situations related to the phobic object. This belief, and its subsequent avoidance, will prevent the patient from acquiring new information disaffirming the catastrophic scenario. Consequently, it prevents the patient from getting new experiences proving the harmless nature of the situation (Öst, 2013).

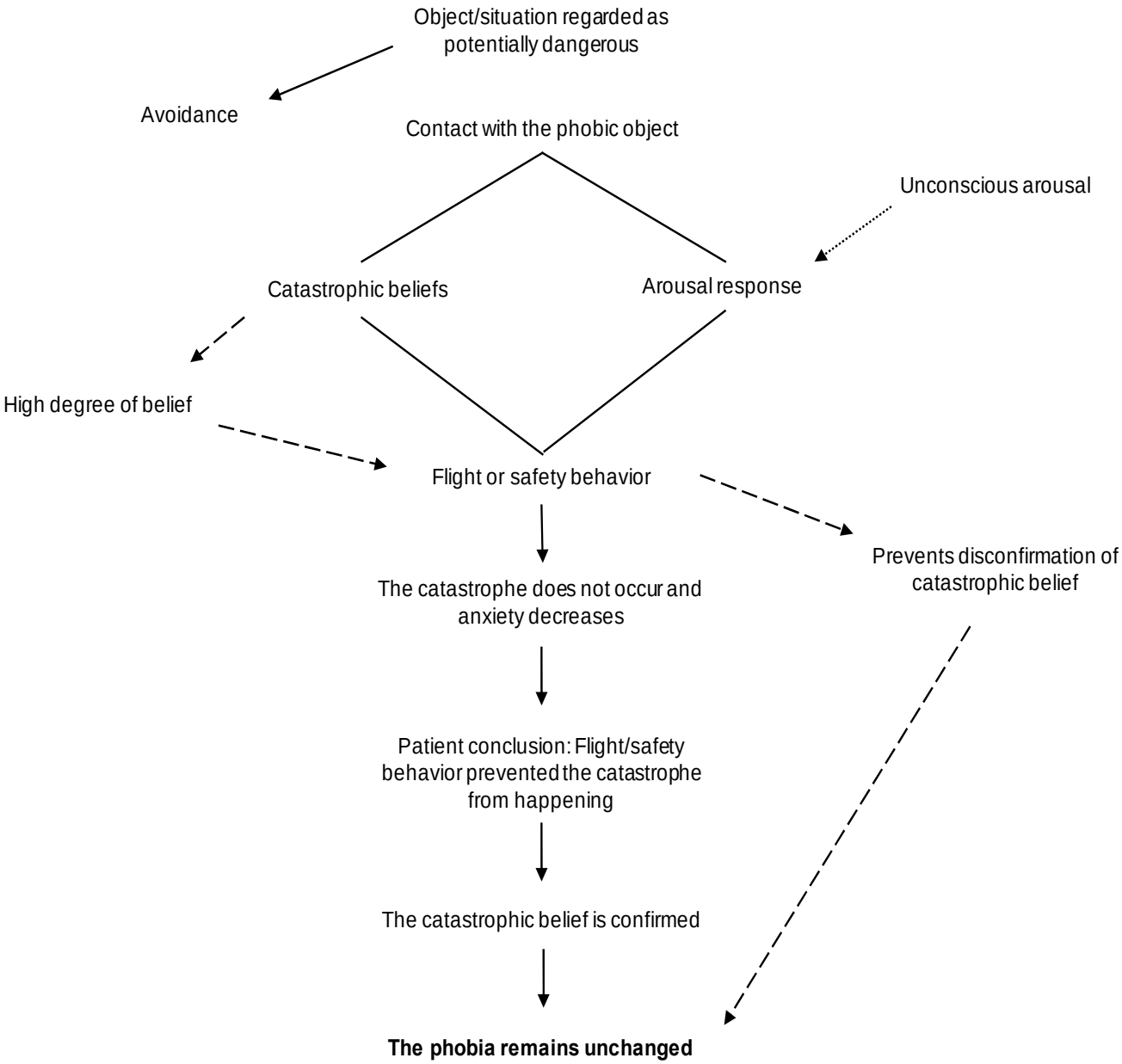

Figure 1. Model describing how specific phobia develops and is sustained (Öst, 2013).

Although specific phobia is impairing, there are fortunately effective treatments available. Cognitive Behavioral Therapy (CBT) with in vivo exposure is the preferable treatment option for specific phobia (Kaczkurkin & Foa, 2015). The one-session treatment (OST; Öst, 1987), a CBT treatment including exposure, can be considered the treatment of choice among CBT-treatments due to its significant positive effects on self-report measures, behavioral measures as well as having cost-beneficial advantages (Ollendick & Davis, 2013; Zlomke & Davis, 2008). The OST consists of three hours of graded exposure to approach the phobic stimulus, combined with patient-therapist teamwork to test catastrophic beliefs (Zlomke & Davis, 2008). A meta-analysis examining the efficacy of the exposure treatment compared to control conditions (both passive and active) show that exposure-based treatment for specific phobia produced larger effect-sizes at post-treatment than no treatment (d = 1.05), placebo (d = 0.48) and cognitive therapy (d = 0.44). Furthermore, among the different treatments classified as exposure, in vivo exposure specifically produced a larger positive treatment effect at post-treatment than did imaginative exposure, Eye Movement Desensitization and Reprocessing (EMDR), VR exposure therapy and guided mastery. However, these effects were not significant at follow-up measurement (Wolitzky-Taylor, Horowitz, Powers, & Telch, 2008).

Despite being a highly effective treatment, some limits regarding the benefits from exposure therapy for spider phobia have been raised in the literature. For instance, lack of accessibility to evidence-based treatments (Stinson et al., 2007), maintaining treatment stimuli and reluctance from patients to participate in exposure therapy (Issakidis & Andrews, 2004) are some of the current limitations proposed. When interviewing successfully treated spider phobic patients, it was found that 90 % of spider phobic patients would not have agreed to treatment if they had known what the treatment would entail beforehand (i.e. to face a real spider) (Thompson, Ollendick, & Öst, 2012).

As a way to meet the above mentioned issues associated with exposure therapy, Virtual Reality (VR) has been proposed as a potential instrument that would enable the delivery of exposure therapy to patients in a way that evades the issues related to traditional exposure therapy. VR treatment has been proposed in prior research to be a potentially effective treatment option for specific phobia (Powers & Emmelkamp, 2008). The idea of VR, a live direct or indirect view of a real world environment, was first raised in the beginning of the eighteenth century. Since the first head-mounted display was developed in the 1960's, the technology necessary to realize the idea of VR has been further developed. Today, VR technology delivering a computer-simulated reality is available to the general public (Baus & Bouchard, 2014).

The technology of VR is dependent on the concept of presence. Presence, i.e. feeling present in the environment around oneself, is regarded as a necessary mediator that allows a virtual environment to activate real emotions. The concept of presence is therefore a propsed theory as to why Virtual Reality may work as a treatment option. As Diemer, Alpers, Peperkorn, Shiban and Mühlberger (2015) states: "We can reasonably assume that, when making sense of a VR environment, people apply the same mechanisms to it as they do to everyday reality". Higher immersion, defined as sophistication of simulation or graphics, will lead to an increased perceived presence. Perceived presence and fear appear to be mutually dependent of one another. A certain level of perceived presence has been shown to be necessary for exposure therapy, but increasing the level of perceived presence does not necessarily enhance the effects of exposure therapy (Diemer et al., 2015).

VR exposure therapy (VRET) has been shown to increase heart rate and skin conductance levels in both patients and controls in an experiment conducted on patients with fear of heights and healthy non-phobic controls. This indicates that the VR environment is perceived as real and does evoke fear. When the patients with fear of heights were looking

down, heart rate levels increased more in patients compared to controls (Diemer, Mühlberger, Pauli, & Zwanzger, 2014). In a study examining arousal response in the form of heart rate in non-phobic individuals, the results showed that the heart rate was significantly higher when exposed to a spider stimulus in the VR setting than when exposed to a neutral stimulus (Hamilton, 2015). These findings indicate that VRET has the potential to simulate phobic encounters that are perceived as real. A review of the psychophysiological effects of VRET concluded that the VRET does provoke a fear response (Diemer et al., 2014). Drawing on the results of treatment studies examining the effects of VRET, the level of arousal obtained in a VR exposure session seems to be enough to give positive results regarding phobic fear and high ratings of subjective fear (Diemer, Lohkamp, Mühlberger, & Zwanzger, 2016; Diemer et al., 2014).

Regarding previous clinical research on VRET, a meta-analysis published in 2015 examined the effects of VR exposure and summarized the current status of VRET (Morina, Ijntema, Meyerbröker, & Emmelkamp, 2015). The analysis, including 14 studies done on the treatment of specific phobia, showed that patients with specific phobia performed significantly better post-treatment in a behavioral approach test (BAT) that measures willingness to approach the phobic object, than pre-treatment. Patients also performed significantly better on a BAT post-treatment than did wait-list controls. Regarding long-term effects, no significant differences between the patients treated with VRET and patients treated with traditional in vivo exposure could be shown at follow-up. The authors conclude that there is support for VRET as an effective treatment for specific phobia and that VR exposure actually can produce significant behavior change in real life (Morina, Ijntema, Meyerbröker, & Emmelkamp, 2015). Another review, including five studies examining the effects of VRET for specific phobia, also confirmed long-term treatment gains, showing that effects were maintained over a period of six months to three years, and were comparable to in vivo exposure (Choy et al., 2007).

Research done on the VRET for spider phobia shows similar results as studies examining VRET for specific phobia in general (Parsons & Rizzo, 2008). A study comparing VRET to exposure therapy for the treatment of spider phobia showed no significant difference between the groups regarding their treatment gains. The authors state that both VR exposure and in vivo exposure are efficient methods for treating spider phobia (Michaliszyn, Marchand, Bouchard, Martel, & Poirier-Bisson, 2010).

Exactly which process/processes that mediate positive treatment outcome in the VR setting need further investigation. Whether the habituation process, i.e. the process of arousal and anxiety increase and decrease during an exposure session, is the process mediating outcome in VRET is still unclear (Diemer et al., 2016). It is not yet verified whether habituation processes are active during a VRET session, and if this could potentially explain positive treatment outcomes. However, as mentioned earlier, there is some evidence indicating that VRET produces sufficient fear-activation (Diemer et al., 2014). This would strengthen the hypothesis that habituation, which is dependent of arousal occurring during an exposure session, can explain positive treatment outcomes in VRET, as it does regarding traditional in vivo exposure. When the evidence regarding processes behind successful in vivo exposure overall has been discussed, inhibitory learning has been offered as an alternative to the traditional habituation theory (Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014). One of the components within inhibitory learning stipulate that effective exposure ensures that the patients expectations before treatment are maximally violated, and thus allows the patient to get a new experience contradicting the initial fearful expectation. This process does not necessarily require an arousal and anxiety increase and decrease (Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014). Whether the process of inhibitory

learning could be a viable explanation regarding the positive treatment outcomes in VRET is yet to be investigated.

In summary, VRET has so far proven to be a potentially effective treatment for treating spider phobia. There are in addition some ways in which VRET could be considered beneficial that should be mentioned. For example, Krijn, Emmelkamp, Olafsson and Biemond (2004) mention the cost effectiveness of VRET compared to traditional in vivo exposure. The authors also raise the benefit of being able to, in a flexible manner, adapt to the client's individual needs and the benefit of being able to repeat exposure tasks over and over. Availability and maintenance of stimulus is another obvious factor, which is a common problem when conducting traditional in vivo exposure. VRET could also encourage people, who would otherwise not seek treatment, to do so (Coelho, Waters, Hine, & Wallis, 2009). In addition, it is very beneficial that VR technology enables researchers to acquire high reliability between participants since treatment procedure and treatment parameters can be held constant (Krijn et al., 2004).

Although VRET has shown promising results in prior studies, there are some aspects that still need to be addressed before establishing the treatment as a formal treatment option for specific phobias. Even though previous studies have shown that VRET can be a potential treatment option to be used by phobic patients afraid of spiders, many of these studies have been of poor quality (McCann et al., 2014). Reviews of the current literature show that more randomized controlled studies of good quality are necessary to establish VRET as a formal treatment option in a healthcare setting. Another aspect that has been mentioned in the literature as a potential issue with VRET is motion sickness experienced by users of the equipment (Krijn et al., 2004). With the development of better VR technology, with less intrusive equipment, the risk of nausea has been minimized (Davis, Nesbitt, & Nalivaiko, 2015). Despite this, the risk of nausea may be something worth monitoring in studies examining the effects of VRET.

Previous research has highlighted the importance of finding out what treatment works for whom. This is usually studied by investigating potential predictors of treatment outcomes (Andersson, Carlbring, & Grimlund, 2008). A predictor variable is, as its name implies, an independent variable used in statistical modeling to predict the value on a dependent variable (the criterion variable). However, it is important to have in mind that, unlike experimental research, research on predictor variables is strictly correlational, i.e. no conclusions about causation can be made (Veličković, 2015).

As mentioned earlier, although effectiveness is high, the availability of traditional CBT with in vivo exposure for the treatment of specific phobia is limited. Therefore, an important research question is to find out what, if any, patient characteristics can predict successful treatment outcome with VRET for specific phobias. Potential predictors would aid clinicians in utilizing healthcare resources more efficiently. This would be in accordance to the stepped care model, i.e. the concept of organizing treatment interventions so that the most effective, yet least resource intensive, treatment is delivered first (Bower & Gilbody, 2005).

Although potential predictors worth investigating have been identified, studies investigating these predictors in anxiety disorders are lacking (Wolitzky-Taylor, Arch, Rosenfield, & Craske, 2012). Likewise, Meyerbröker and Emmelkamp (2010) argue that more research is needed regarding treatment outcome predictors, especially in VRET. Potential predictors proposed by previous research regarding specific phobia include: age, onset, duration, initial level of anxiety and depression, overall phobia, initial level of complaints, credibility and expectations from treatment, way of acquisition, family prevalence of the same phobia, heart rate and systolic blood pressure and diastolic blood pressure at pre-treatment behavior test (Hellström & Öst, 1996). Among these potential predictors, outcome expectancy, the patient's beliefs about whether they will benefit from

treatment or not, has been identified as an important predictor for treatment gains in previous research examining predictors of treatment outcome (Price, Anderson, Henrich, & Rothbaum, 2008). However, Price et al. (2008) points out that most of the research that has highlighted the importance of outcome expectancy was conducted over two decades ago, and that more recent research on outcome expectancy for the treatment of anxiety disorders have had mixed results. Another predictor that has previously been shown to be important for treatment gains is initial level of phobic fear or initial level of impairment due to the disorder (Nordgreen et al., 2015). The potential predictors outcome expectancy and initial impairment of phobia are therefore both relevant to study further in a VR setting.

In summary, VRET is a potentially cost-effective and accessible treatment for spider phobia, but more research is needed on predictors regarding individual characteristics that could potentially be relevant for the clinician making a decision which type of treatment to give each patient. This could, for example, involve what type of patient that would seem to benefit sufficiently from VRET and what type of patient that may require the traditional OST-approach to be able to benefit from the treatment.

The purpose of this study is to examine if the scores on the self-rating scales (measuring initial level of phobic fear) Spider Phobia Questionnaire (SPQ; Klorman, Weerts, Hastings, Melamed, & Lang, 1974), Fear of Spiders Questionnarie (FSQ; Szymanski & O'Donohue, 1995) as well as the Treatment Credibility Scale (TCS; (Borkovec & Nau, 1972) can predict scoring on the primary outcome measure (Behavioral Approach Test, BAT), in the two separate treatment conditions OST and VR OST. We hypothesize that:

1) the pre-treatment total score on the SPQ and the FSQ predicts treatment outcome on the BAT in both the OST and VR OST,
2) TCS at pre-treatment predicts treatment outcome on the BAT in both the OST and VR OST.

Method

Participants

Sample

For this study, a sample consisting of self-selected participants (n = 100) was collected. Participants were recruited through different media (i.e. through public television, newspapers, online forums and magazines) in order to reach a diverse sample. Potential participants were directed to the study website. The demographic variables of the collected sample are described in Table 1.

Inclusion criteria

To be included in the study, the participants had to fulfill the DSM-5 criteria for specific phobia of spiders, measured with the Structured Clinical Interview for DSM disorders (SCID-I/P; First, M. B., Spitzer, R.L, Gibbon M., and Williams, 2002). Furthermore, participants had to be at least 18 years of age, have Swedish residency and fluency, being able to attend all measurements carried out at four occasions and had to score less than 10 points on the BAT.

Exclusion criteria

The participants were excluded from the study if there were indications of suicidal ideation, indications of other serious mental health disorders (e.g. substance abuse, bipolar disorder, psychosis), if there was a lack of stereoscopic vision or balance problems (that would make it hard for the participant to participate in the VR environment) as well as if participants had an ongoing psychotherapy or pharmacotherapy (unless the participant had had a stable dosage in their pharmacotherapy for the past three months).

Table 1. Sample demographics

		OST (*n* = 50)	VR (*n* = 50)	Total (*n* = 100)
Age	*M*	34.04	34.06	34.05
	SD	9.85	10.92	10.35
Gender	*Male*	8 (16 %)	8 (16 %)	16 (16 %)
	Female	41 (82 %)	42 (84 %)	83 (83 %)
	Other	1 (2 %)	0	1 (1 %)
Level of education	*Elementary school*	3 (6 %)	0	3 (3 %)
	High school	18 (36 %)	15 (30 %)	33 (33 %)
	University	29 (58 %)	35 (70 %)	64 (64 %)
Occupation	*Employed*	39 (78 %)	35 (70 %)	74 (74 %)
	Parental leave	1 (2 %)	3 (6 %)	4 (4 %)
	Job seeker	2 (4 %)	1 (2 %)	3 (3 %)
	Student	8 (16 %)	10 (20 %)	18 (18 %)
	Retiree	0	1 (2 %)	1 (1 %)
Psychotherapy	*Current or past*	13 (26 %)	10 (20 %)	23 (23 %)
	No	37 (74 %)	40 (80 %)	77 (77 %)
Pharmacotherapy	*No earlier use*	44 (88 %)	46 (92 %)	90 (90 %)
	Past use	6 (12 %)	0	6 (6 %)
	Current use	0	4 (8 %)	4 (4 %)

Note. OST = One-Session Treatment, VR = Virtual Reality.

Materials and Apparatus

Primary outcome measure

In this study the BAT (Öst, Salkovskis, & Hellström, 1991) served as the main outcome measure. The BAT featured 13 graded steps (scored 0-12) and included a medium sized Swedish spider to test for generalization to real world effects of both treatments. The BAT starts with the participant standing outside a room where there is a spider inside. The spider is located in a transparent plastic container (approximately 20 cm x 30 cm x 20 cm) with a transparent lid on a table in the far end of the room. The participant is instructed to get as close to the spider as he/she can manage, with the aim of picking up the spider and holding it in their hands for 20 seconds. They are encouraged to do their best, but are also told that they can abort the test at any time. The score obtained on the BAT is based on how close to the spider the participant manages to get without therapist involvement. Transparent pieces of tape, corresponding to BAT-scoring, were put on the floor to ensure all therapists used the same distances for scoring. See Table 2 for details on scoring.

Table 2. Point system in the Behavioral Approach Test (BAT)

Score	*BAT-task difficulty level*
0	Will not enter room.
1	Enters room but stops before 1/5 of the distance to the container is reached.
2	Stops before 2/5 of the distance to the container is reached.
3	Stops before 3/5 of the distance to the container is reached.
4	Stops before 4/5 of the distance to the container is reached.
5	Stops before the whole distance to the container is reached.
6	Stops in front of table with container.
7	Touches the container.
8	Takes off the lid of the container.
9	Places hand in container.
10	Pokes spider with at least one finger.
11	Holds spider in hands for < 20 seconds.
12	Holds spider in hands for ≥ 20 seconds.

Secondary measures

All the secondary measurements used in the study are described. Two self-rating scales measuring spider phobia symptoms were used, namely the Spider Phobia Questionnaire (SPQ; Klorman, Weerts, Hastings, Melamed, & Lang, 1974) and the Fear of Spiders Questionnaire (FSQ; Szymanski & O'Donohue, 1995). The SPQ is a 31-item scale consisting of 31 statements that can be answered either true or false (a dichotomous scale). A high score indicates a greater phobic fear of spiders (with total scores ranging from 0-31). The psychometric properties of this scale can be considered adequate, with stable results with regard to reliability and validity. The scale has a sufficient internal consistency with a Cronbach's $\alpha = 0.62$-0.90 (Muris & Merckelbach, 1996). The FSQ consists of 18 phobic-related statements regarding spiders (with total scores ranging from 0-108). Participants stated how much they agreed to the statement using a Likert scale ranging from 0-6, where a higher number indicates a higher degree of of spider phobia severity. The questionnaire has been shown to have good test-retest reliability, excellent internal consistency with a Cronbach's $\alpha = 0.92$, and a significant convergence validity demonstrated by significant correlations with both the SPQ and the BAT (Szymanski & O'Donohue, 1995).

Treatment credibility was assessed using the self-rating scale Treatment Credibility Scale (TCS; (Borkovec & Nau, 1972). The TCS is a 5-item scale. Participants rated the questions on a scale of 1-10 (with total scores ranging from 0-50). A higher score indicates a greater belief that the treatment would have positive effects. In this study, the scale was administered twice for each participant with ratings made for both treatment conditions. The participant asked to rate the 5 items each for the OST and VR OST (leaving a rating for TCS-OST and a rating for TCS-VR OST), and the order in which the two scales appeared in the self-rating scales battery was randomized between participants (as to avoid bias regarding order of appearance). This questionnaire is not validated in the Swedish language and therefore it is hard to state psychometric properties regarding the scale used in this thesis. See Appendix A for presentation of the TCS scales.

Also in this study, generic anxiety symptoms were measured with the Generalized Anxiety Disorder 7-item scale (GAD-7; Spitzer, Kroenke, Williams, & Lowe, 2006), depressive symptoms were measured with the Patient Health Questionnaire 9-item scale (PHQ-9; Kroenke, Spitzer, & Williams, 2001) and subjective quality of life was measured

with the Brunnsviken Brief Quality of Life Scale (BBQ; Lindner et al., 2016). The presence of a diagnosis for specific phobia for spiders measured with SCID-I/P (SCID-I/P; First, M. B., Spitzer, R.L, Gibbon M., and Williams, 2002) was used as a secondary outcome measure. The data from the secondary outcome measures that were not of importance for our hypotheses were not used in this thesis (i.e. the GAD-7, the PHQ-9, the BBQ and the SCID-I/P).

Electrical brain activity was measured with electroencephalography (EEG). Recordings were obtained while participants were asked to look at pictures with negative valence, neutral valence, positive valence, and pictures portraying spiders. After looking at the pictures, the participants were asked to rate a representative sample of them on scales of liking and arousal. The EEG was always conducted after the other measurements in an EEG-lab at the study site. The data collected in the EEG measurement is not analyzed or presented in this thesis.

Interventions

The traditional OST consisted of gradual exposure to the phobic stimuli, modeled by the therapist (i.e. the therapist always begins by demonstrating how to interact with the spider in a non-phobic manner). During the treatment, the participant was asked to rate their subjective unit of distress (SUD) on a scale from 0-100. Each OST lasted for three hours, starting with psychoeducation about specific phobia and anxiety, and rationale for the treatment. Three different sizes of spiders were used in this study: small (0.5-1.5 cm), medium (1.5-2.5 cm) and large (> 2.5 cm). All size measurements are including the legs of the spider. For each spider size, the therapist modeled four behaviors that the participant should perform before moving on to the next behavior. When all four behaviors for a certain spider size were completed, the same procedure was repeated with a larger size. The four behaviors were: capture the spider with a glass and postcard, poke the spider with a finger, hold the spider in one's hands and let the spider crawl on one's body. After the three-hour session, participants were given information about maintaining progress and the importance of continuing to practice exposure on their own.

The VR OST consisted of an application programmed to guide the participants through a number of zones where they complete gamified tasks which gradually increased in intensity and level of realism. See Figure 2 for examples of spider stimuli. The application was delivered through a head-mounted display with headphones. During the treatment, the participant was asked to rate their SUD on a scale from 0-100.

Figure 2. Pictures of three different spiders presented in the VR OST application (Vimse).

The virtual environments were delivered to the participant by combining a Samsung Gear VR head-mounted display (HMD) with a Samsung Galaxy device. Two different setups were used. The Gear VR First Innovator Edition was paired with Samsung Galaxy Note 4,

whereas the Gear VR Second Innovator Edition was paired with Samsung Galaxy S6. Both Galaxy devices were running Android 5.0.1.

Procedure

The present study has been registered in the Clinicaltrials.gov database (NCT02533310) and has received ethical approval from the Stockholm Regional Ethical Review Board (Dnr: 472-31). The advertisements marketing the study were posted in online forums, newspapers and magazines, and shown on national television. All advertisements referred to the study website (http://www.vimse.se) where participants could find information and self-register for participation. See Figure 3 for study flow chart describing participant flow through the study. Participants registered at the website and acquired an account in the platform iTerapi (http://www.iterapi.se). All participants provided a written and informed consent prior to participation. Information provided to the participant included the purpose of the research, their right to decline to participate and withdraw at any time, confidentiality and contact information. Besides providing demographic data, participants also filled out the FSQ, the SPQ, the GAD-7, the PHQ-9 and the BBQ. Participants who fulfilled the online screening criteria were invited to participate in the pre-measurement conducted at the study site.

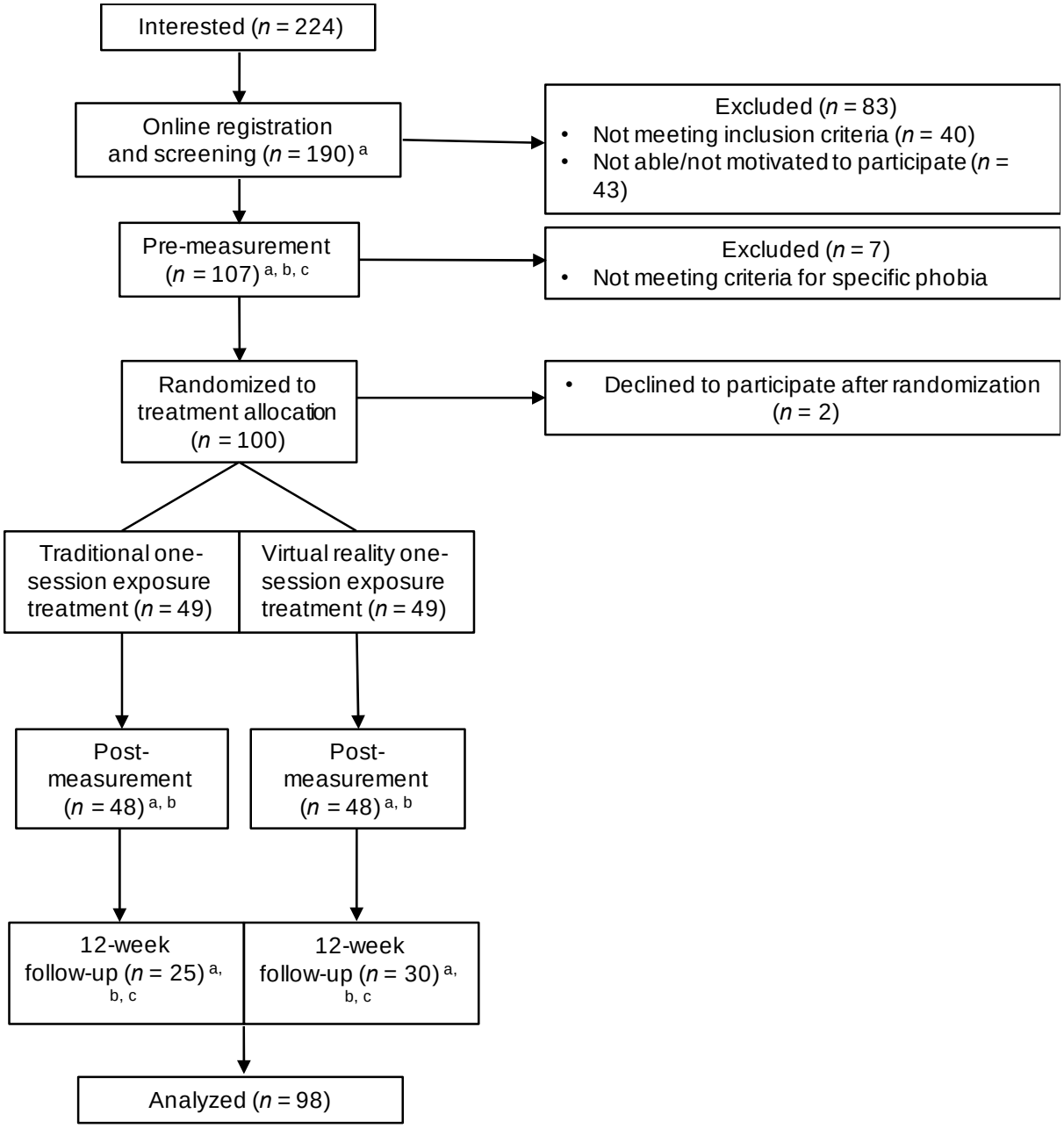

Figure 3. CONSORT flow chart.
Note. [a] Self-rating scales, [b] BAT, [c] Diagnostic interview. At the time of this thesis, the 12-week follow-up data collection was still in progress.

Measurements and treatments were carried out by a psychologist in clinical supervised practice, three psychotherapy students and eight psychology master students. All therapists and students involved in the OST treatment were trained in a workshop format by a licensed psychologist/psychotherapist in the method and approach of the OST. The psychologist/psychotherapist responsible for the training in the OST was also available on at least one occasion for supervision and coaching for all therapists and students involved in the delivery of the OST. Demonstration and training in the procedure of the VR OST was provided for all involved in treatment by the company responsible for developing the VR OST application Vimse, namely Mimerse.

The pre-measurement consisted of further screening and baseline measuring, conducted by the therapists/students involved in the study. During pre-measurement, the participants were first interviewed with the SCID-I/P to confirm phobia diagnosis. They were then asked to answer a battery of questionnaires, including the SPQ and the FSQ as well as the TCS. Next, the BAT was performed. Lastly, the participants underwent the EEG measurement, conducted by EEG-technicians in a separate facility. Pre-measurements took about one hour to complete.

After pre-measurement was completed, participants were randomized to either OST or VR OST, and randomized to a therapist who was blinded for the individual carrying out the pre-measurement. Except for treatment condition, all other variables were equal for all participants. Treatment sessions were conducted about 7 days after pre-measurement. During the treatment session, the participant was first informed on what type of treatment they had been randomized to, if they had not been told earlier by telephone when booking the treatment session. If randomized to OST, the therapist simply initiated the treatment as described previously. If randomized to VR OST, the participant was first introduced to the technicalities of the VR equipment by the therapist/student. They then proceeded to go through the VR OST by themselves (self-administered) with the therapist/student acting as technician providing assistance with the equipment if needed. Both groups of treatment were provided with a relapse prevention strategy/plan to help them maintain their treatment results. The treatment sessions took approximately three hours to complete. See section on treatment conditions in introduction for more detailed information regarding the interventions.

Post-measurements were conducted approximately 7 days after the treatment sessions were carried out. Post-measurements were executed by a therapist different from the one who performed the treatment session and who was blinded for therapist responsible for treatment session. To avoid biased BAT-scoring, participants were instructed not to tell the therapist what treatment he/she had undergone. The post-measurements followed the same protocol as the pre-measurements, except for the addition of a self-rating scale measuring negative effects of psychotherapy (Rozental et al., 2014) and the removal of the SCID-I/P diagnostic interview. Post-measurements took approximately one hour to complete.

Follow-up measurements of the participants were conducted around 12 weeks after the post-measurements. The procedure followed the same routine as the pre-measurement, with a therapist blinded to treatment condition carrying out the measurement.

An additional follow-up measurement is planned to be carried out about one year after post-measurement. The procedure will follow the same routine as the pre-measurement and 12-week follow-up measurement.

Data Reduction and Statistical Analyses

Data reduction

The raw dataset was obtained, containing all variables measured in the study. The variables of interest, with regards to this thesis hypotheses, were sorted out and formatted in a new file. This file, formatted in the wide form (the standard form in all statistical software), was then re-formatted into the long form as to enable the later statistical analyses. The conditions (treatment groups) were binary coded (OST = 0, VR = 1), and the occasions (measurement time points) were numerically coded (pre-measurement = 0, post-measurement = 1, 12-week follow-up = 2) in line with recommendations made by Hesser (2015).

Power analysis

There is no accepted standard for calculating power a priori in a growth curve multilevel model (Castelloe & O'Brien, 2001). If calculating a power analysis for a mixed ANOVA, theoretically, a between-group effect size at Cohens $d = 0.57$ could be detected when the alpha-level is set to $p < .05$ and a power of 80 % if 100 participants (with 50 in each

group) would be included. Generally when applying a multilevel model, power is considered to be larger than when applying an analysis of variance (Fan, 2003).

Statistical analyses

Data was analyzed using R v 3.2.5 for Windows and Mac (R Core Team, 2016). An independent samples t-test was used to assess if any significant differences in predictor means between treatment groups existed at pre-measurement (assessed as a precaution). Furthermore, a growth curve model (multilevel model analysis) was developed to test the influence of the hypothesized predictors (FSQ, SPQ and TCS) on the primary outcome measure (BAT) at all occasions, i.e. pre-measurement, post-measurement and 12-week follow-up. The growth curve model analyses in R were computed using the nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2016).The growth curve model was analyzed including all participants, regardless of whether they completed the study or not, in accordance with the intent-to-treat principles recommended by Hesser (2015).

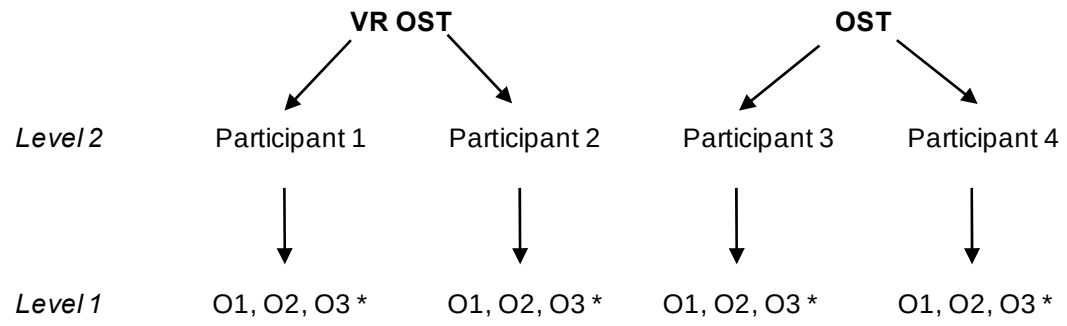

Figure 4. Hierarchical multilevel model with occasions nested within individuals.
Note. * O = Measurement occasion.

The hierarchical multilevel model was developed based on recommendations by Hesser (2015), Peugh (2010) and Steele (2014). A summary of the model is found in Figure 4. First, a simple null model (unconditional model) was formulated in an equation as the base for further analysis.

Level 1: $Y_{ij} = \beta_{0j} + R_{ij}$
Level 2: $\beta_{0j} = \gamma_{00} + U_{0j}$

In the equation, γ_0 is defined as the overall intercept in the model (with β_{0j} being the intercept for individual $_j$), R_{ij} defined as the within-individual variance in Y and U_{0j} defined as individual-specific random effects.

Second, a growth model (unconditional growth model) equation was formulated adding occasion (measurement time point) as an explanatory variable to the primary outcome measure.

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j}OCCASION_{ij} + R_{ij}$
Level 2: $\beta_{0i} = \gamma_{00} + U_{0j}$
$\beta_{1j} = \gamma_{10} + U_{1j}$

In the equation, β_1 is defined as the slope of the regression or growth rate. The model included occasion as a fixed effect.

Third, a growth model (conditional growth model) equation was formed adding condition (treatment group) as another explanatory variable together with occasion. This model featured a random slope and random intercept.

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j}\text{OCCASION}_{ij} + R_{ij}$$
$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}\text{CONDITION}_j + U_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}\text{CONDITION}_j + U_{1j}$$

In this model, condition was added as a fixed effect together with occasion. Furthermore, the predictor variables were added to the conditional growth curve model in a later stage.

Results

The independent samples t-test used to assess differences in predictor means between the two treatment groups at pre-measurement showed that no significant differences existed regarding any of the predictors. Means and standard deviations of the variables of interest at all measurement occasions are presented in Table 3. Pre-analysis diagnostics were made to assess the assumptions made in a multilevel model. These diagnostics showed that the assumption of normally distributed residuals was met in this data set. However, the assumption of homogeneity of variances was not fully met. Fortunately, the use of a repeated measures multilevel model overrides the assumption of homogeneity of variances (Cohen, West, & Aiken, 2013).

Due to the possibility of ceiling effects for the OST group's results on the BAT, as apparent by Figure 5, a Bartlett's test was conducted ($p > .05$) to test examine the possibility of unequal variances between OST and VR. No further action regarding the homogeneity was taken due to the test being non-significant.

Missing data at post-treatment and 12-week follow-up was treated as missing at random in the analyses due to the fact that there was no significant difference in the attrition rate between the two conditions. Attrition rates for the two conditions at post-treatment were 7 % ($n = 3$ for OST and $n = 4$ for VR). While this thesis was still being written, the 12-week follow up data collection was still in progress. At the time for this thesis, 55 % of participants have been successfully contacted ($n = 55$), of which $n = 30$ (60 %) are in the VR group, and $n = 25$ (50 %) are in the OST group.

Table 3. Mean and standard deviation of primary outcome value and predictor values at all three occasions

	OST			VR		
	Pre-measurement	*Post-measurement*	*12 week follow-up*	*Pre-measurement*	*Post-measurement*	*12 week follow-up*
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
BAT	5.66 (2.47)	10.70 (1.68)	10.92 (1.29)	4.76 (2.71)	8.50 (2.89)	9.00 (2.60)
FSQ	80.01 (13.91)	36.59 (17.81)	35.29 (21.68)	77.82 (15.26)	52.35 (22.45)	45.03 (23.25)
SPQ	17.76 (2.06)	12.55 (2.99)	13.12 (3.46)	17.44 (2.38)	13.85 (3.55)	13.70 (4.64)
TCS - OST	39.06 (7.14)	-	-	36.22 (10.18)	-	-
TCS - VR	34.78 (7.57)	-	-	31.90 (9.78)	-	-

Note. OST = One-Session Treatment, VR = Virtual Reality.

The growth curve model fit was assessed using a likelihood-ratio test. Table 4 shows estimated coefficients, residuals, goodness of fit and tests of significance. The results showed that there was a significant effect of occasion and condition on the primary outcome measure (BAT), entailing that occasion alone could explain 60 % of variance in BAT, while adding type of condition increased this explanatory value to 64 %. Because there was a significant effect (Cohen's $d = 1.10$ on difference in BAT between groups at post-treatment) of condition (i.e. type of treatment) on the BAT, data from each treatment group was divided into two data files and analyzed separately in subsequent analyses.

Table 4. Results from a linear growth model of predictors of treatment outcome in both treatment groups

	Unconditional model		Unconditional growth model		Conditional growth model	
	Estimate (SE)	DF	Estimate (SE)	DF	Estimate (SE)	DF
Fixed effects						
Intercept	7.87 (0.22)*	148	5.76 (0.25)*	147	6.32 (0.35)***	146
Occasion	-		2.62 (0.17)*	147	2.92 (0.24)***	146
Condition	-		-		-1.13 (0.49)*	98
Interaction	-		-		-0.58 (0.34)	146
Random effects						
Residual	9.88		3.91		3.68	
Intercept	0.82		2.80		1.64	
	Deviance		Deviance		Deviance	
Model summary						
Goodness of fit	645.17		571.26***		557.80***	

Note. Unconditional growth model does not include condition; the conditional model includes condition. Condition is a binary coded variable (1 = VR OST, 0 = OST).

* $p < .05$. ** $p < .01$. *** $p < .001$.

In the following analysis, growth curve models investigating the effects of occasion on the BAT was made with data from only the VR OST condition. The predictors FSQ, SPQ, TCS rating for the OST (TCS-OST) and TCS rating for the VR (TCS-VR) were added separately for each model as random effects. Table 5 shows estimated coefficients, residuals, goodness of fit and tests of significance. Percent of explained variance for the predictors with significant effects on the primary outcome measure was 9 % for FSQ, 17.7 % for SPQ, and 0 % for TCS-OST. TCS-VR had a non-significant effect in the model and explained 0 % of the variance in the BAT. Figure 5A-D shows the relationship between predictor variables at pre-treatment and the primary outcome measure (BAT) at post-treatment.

Table 5. Results from a linear growth model of predictors of treatment outcome in the VR OST condition

	Unconditional growth model		Conditional growth model with predictor FSQ		Conditional growth model with predictor SPQ		Conditional growth model with predictor TCS (rating for VR OST)		Conditional growth model with predictor TCS (rating for OST)	
	Estimate (SE)	DF	Estimate (SE)	DF	Estimate (SE)	DF	Estimate (SE)	DF	Estimate (SE)	DF
Fixed effects										
Intercept	5.17 (0.36)	75	10.09 (0.78)***	74	10.91 (1.14)***	74	3.17 (1.23)**	74	1.50 (1.25)	74
Occasion	2.40 (0.21)	75	1.12 (0.27)***	74	1.44 (0.26)***	74	2.87 (0.77)***	74	2.31 (0.77)**	74
Predictor	-	-	-0.06 (0.01)***	74	-0.33 (0.06)***	74	0.06 (0.04)	48	0.10 (0.03)**	48
Random effects										
Residual	3.00		2.72		2.47		2.99		3.01	
Intercept	3.88		1.74		3.45		3.53		2.83	
	Deviance		Deviance		Deviance		Deviance		Deviance	
Model summary										
Goodness of fit	283.97		265.49***		271.30***		282.50		277.87**	

Note. Unconditional growth model does not include condition; the conditional model includes condition. Scores on the primary outcome measure and predictors indicate the mean item response.

* $p < .05$. ** $p < .01$. *** $p < .001$.

The same analyses executed in the VR OST condition were made with data only from the OST condition. Table 6 shows that mean scores on the FSQ and SPQ have a significant effect on the BAT over all occasions. Adding FSQ as a predictor explained 75 % of variance in the BAT, while adding the SPQ as a predictor explained 42 % of variance in the BAT. The TCS-VR and TCS-OST explained 0 % of the variance in the BAT, and their coefficients were non-significant. Figure 5A-D shows the relationship between predictor variables at pre-treatment and the primary outcome measure (BAT) at post-treatment.

Table 6. Results from a linear growth model of predictors of treatment outcome in the OST condition

	Unconditional growth model		Conditional growth model with predictor FSQ		Conditional growth model with predictor SPQ		Conditional growth model with predictor TCS (rating for VR OST)		Conditional growth model with predictor TCS (rating for OST)	
	Estimate (SE)	DF	Estimate (SE)	DF	Estimate (SE)	DF	Estimate (SE)	DF	Estimate (SE)	DF
Fixed effects										
Intercept	6.32 (0.30)	71	11.86 (0.63)***	70	12.37 (1.07)***	70	5.33 (1.06)***	70	6.68 (1.26)***	70
Occasion	2.96 (0.27)	71	1.22 (0.30)***	70	1.87 (0.28)***	70	2.95 (0.27)***	70	2.96 (0.27)***	70
Predictor	-	-	-0.08 (0.01)***	70	-0.36 (0.06)***	70	0.03 (0.03)	48	-0.01 (0.03)	45
Random effects										
Residual	4.84		1.23		2.82		4.88		4.85	
Intercept	0.31		3.96		2.66		0.23		0.31	
	Deviance		Deviance		Deviance		Deviance		Deviance	
Model summary										
Goodness of fit	273.03		245.49***		258.24***		272.54		272.99	

Note. Unconditional growth model does not include condition; the conditional model includes condition. Scores on the primary outcome measure and predictors indicate the mean item response.

* $p < .05$. ** $p < .01$. *** $p < .001$.

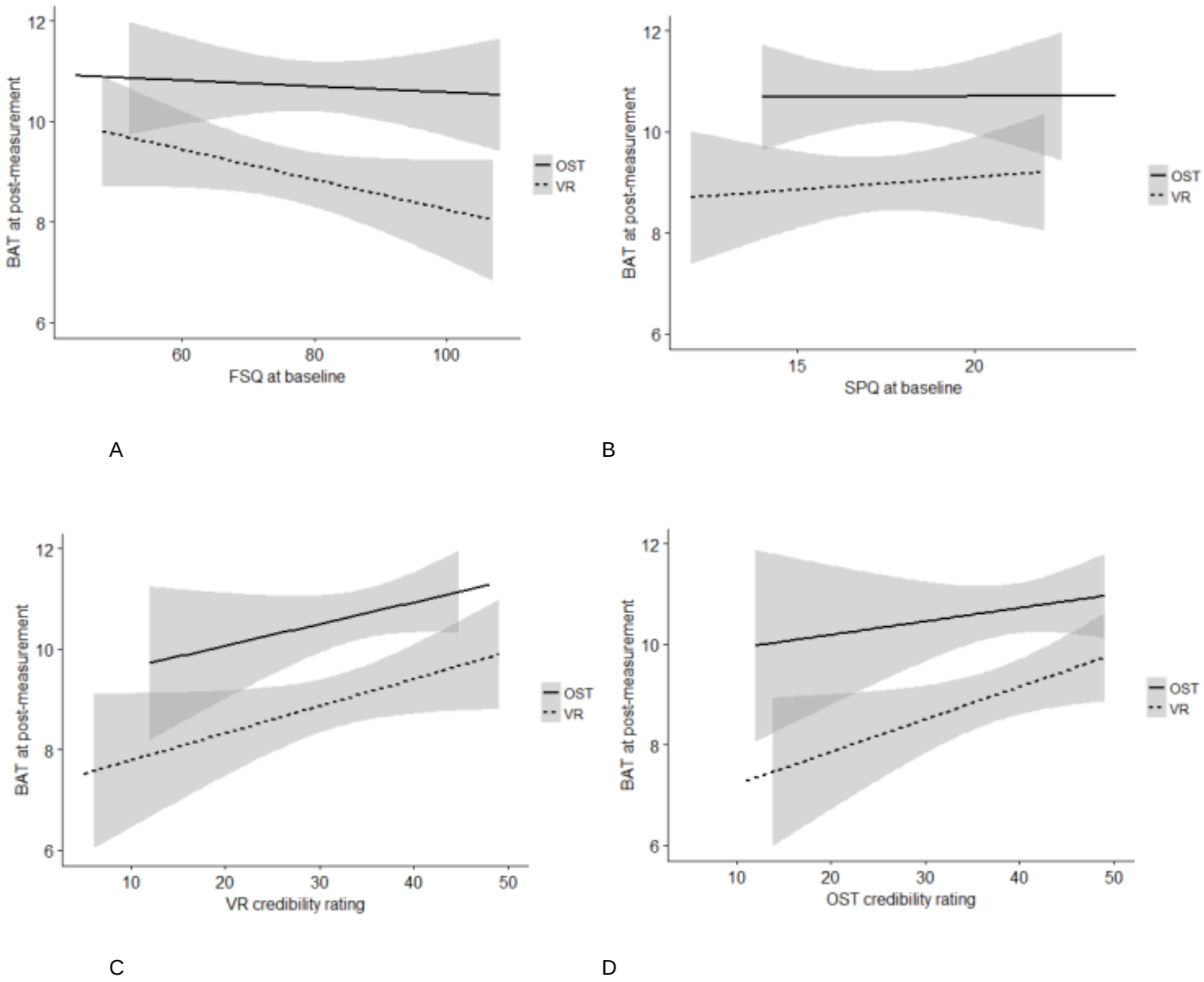

Figure 5 A-D. Relationship between the primary outcome measure (BAT) at post-treatment and baseline measures of the predictors FSQ (Figure 5 A), SPQ (Figure 5 B), TCS rating of VR (Figure 5 C) and TCS rating of OST (Figure 5 D). The figures illustrate how scores on the BAT at post-measurement relate to the predictors investigated in this study. The plots display the linear regression between the primary outcome measure and the predictor (line in black) and the confidence interval stated at the 95 % confidence level (area in grey). The relationships are based on the models presented in Table 5 and 6.

Discussion

The aim of this study was to test the two hypotheses stating that scoring on the self-rating scales SPQ and FSQ (measuring initial impairment of phobia), as well as whether treatment credibility for the OST and VR treatment could predict treatment gains measured by the primary outcome measure (BAT), in the two separate conditions OST and VR OST. The results suggest that initial impairment of phobia is a predictor of treatment outcome, and

the most important predictor to take into account of the predictors analyzed. The predictor treatment credibility showed more ambiguous effects on the treatment outcome. The primary results in this study, that initial impairment of phobia affects treatment outcome, is in line with previous research done in the field of anxiety treatment research (Hellström & Öst, 1996; Scholing & Emmelkamp, 1999). The fact that treatment credibility showed effects that were harder to interpret, and effects that were sometimes not significant, is also in line with previous research done on treatment credibility as a predictor of treatment outcome in CBT treatments (Greenberg, Constantino, & Bruce, 2006). Greenberg, Constantino & Bruce (2006) states that the status of treatment credibility as a factor important to treatment outcome has varied historically. The ambiguous results regarding treatment credibility could have several explanations. One potential explanation is that treatment credibility has been defined and measured in different ways in previous research, making it hard to come to any significant conclusions about whether the influence of treatment credibility is important or not (Borkovec & Nau, 1972; Greenberg et al., 2006). Another possible explanation to the varying results regarding the effect of treatment credibility on treatment outcome is the treatment results in the OST group as measured by the BAT. The BAT scores in the OST condition showed a clear ceiling effect for the post-measurement results, leaving little variation to explain.

In further detail, the results in the VR OST condition were consistent with our hypotheses meaning that the potential predictors SPQ, FSQ and the TCS rating of OST can be considered predictors of treatment outcome. However, the analyses showed that the TCS rating of VR OST could not be considered a predictor of treatment outcome. The finding that credibility for the OST, but not for the VR OST, predicts treatment outcome in the VR condition calls into question the predictive quality of both scales. One of the possible explanations for the findings regarding the predictor TCS may be that the TCS rating for OST better measures the credibility of CBT in general, and thus tells something about the participant's attitude towards behavioral therapy (Borkovec & Nau, 1972). Another possibility is that the significant effect of TCS rating for OST was just a coincidence, or that the sample size was too small to detect an effect of the TCS rating for VR OST.

Far more distinctive are the results obtained of predicting treatment outcome by initial impairment of phobia. These results show that lower initial impairment of phobia is associated with better treatment outcome especially in the VR OST condition, but also in the OST condition. The effect is more prominent when assessing the predictive value of the FSQ, which suggests the FSQ is a better predictor of treatment outcome than the SPQ. Part of this may be explained by the fact that the FSQ is a continuous measure as opposed to the SPQ which consists of binary items. The data in this study show that the variances are more homogenous in the SPQ which could mean the SPQ is not as sensitive to varying degrees of phobia as the FSQ. This is in line with research done on the psychometric properties of the scales, research showing that the FSQ is more sensitive to measuring fear in the non-phobic range than the SPQ (Muris & Merckelbach, 1996).

The current study had some limitations of importance. As in the majority of research, the methodological set-up and analyses in this study can evaluate potential relationships between the outcome measure and the predictors, but cannot establish a causal link between these. This means that this study cannot prove whether the predictors affect the primary outcome measure directly, or that the effect arises for example via a third variable (Veličković, 2015). This issue of causality is something that unfortunately is difficult, if not impossible, to address in this study, but is worth considering. Due to this study being a double-blind, randomized controlled study, the amount of control over dependent and independent variables is relatively high which for example minimizes allocation bias (Moher et al., 2012).

In this study, the primary outcome measure used was the BAT, a behavioral test which involves approaching a real spider. An argument could be made that this does not evaluate the separate treatment groups in a fair way. This because the VR OST group never faced a real spider during their treatment session, whereas the participants in the OST group faced three real spiders during their treatment session. As mentioned earlier, according to research describing the potential mechanisms of exposure therapy, according to the inhibitory learning theory the experiences of a real encounter with a spider should contribute to new valuable knowledge to rival out previous fearful experiences (Bouton, 1993; Craske, Treanor, Conway, Zbozinek, & Vervliet, 2014). This is something that should be kept in mind while interpreting the results. A more valid way of measuring treatment outcome for participants in a VR exposure therapy without involving a real spider, is something to be developed further as to address this issue. This could be suggested as a future area of investigation. Regarding measurements of initial level of phobia, these are well established instruments (Klorman, Weerts, Hastings, Melamed, & Lang, 1974; Szymanski & O'Donohue, 1995). Particularly the FSQ seems to assess well and be sensitive to the severity of the phobia in the participants in the sample of this study.

Another methodological limitation of this study is the lack of an inactive control group. This means that it is not possible to control for potential outcome of receiving no treatment and weigh this factor into the analysis (Bailey, 2008). Also, a limitation is that this study only analyses data from three measurement occasions. How to calculate and decide the number of occasions required to perform a growth curve model analysis is not straightforward and depends on the nature of the data (Hesser, 2015). According to Hesser (2015), having a greater number of observations is beneficial and provides more intra-individual information. Unfortunately, the one-session format does not in a feasible way provide a possibility to measure individuals at several time points. The power of this study could still be considered as relatively high because of the large sample size and the study design as well as the type of analysis model applied to the data (Curran, Obeidat, & Losardo, 2010).

Something that could also potentially affect the results is that the gender ratio of the participants in the study was skewed, resulting in a majority of female participants. It is though doubtful whether this skewed gender ratio is of importance for the results in this study. However, the generalizability of the results obtained in the study could be discussed, due to that the sample was relatively homogenous and not very diverse in its demographic characteristics.

Despite these limitations, this study adds new valuable information that can have clinical implications. One of the primary findings, namely that initial impairment of phobia predicts outcome in the VR treatment, can potentially be of importance to clinicians treating spider phobia. The participants with less initial impairment of phobia have better treatment outcomes from the VR OST than do those who are more impaired from the start by their phobia. From a cost-effectiveness point of view, the VR OST has many benefits compared to the traditional OST as mentioned earlier in this thesis. To optimize healthcare cost-effectiveness, the stepped care model has been shown to have clinical and economic benefits (Bower & Gilbody, 2005). According to the stepped care model, this means that patients with less symptom load can get treatment more efficiently with briefer interventions that are more accessible. If an individual with a light to moderate spider phobia very easily and cost-effectively can access an effective alternative treatment, such as VR OST, this ought to be a more preferable alternative. This means, and is highlighted by the results of this study, that it could be useful, and in line with the stepped care model, for clinicians to beforehand screen for initial level of phobia in patients that are subject to exposure treatment. The application used in this study, Vimse (now named Itsy) developed by Mimerse, is now available to the public.

Future research should aim towards replicating these results in different settings and with different types of VR treatments. This since this study only assesses the predictive value of initial impairment of phobia specifically regarding the VR OST (Vimse) investigated in this study.

The current study and its results have implications in aspects of everyday life as well. Specific phobia is as aforementioned the second most common mental health disorder and affects almost all of us in some way, either directly or indirectly. Development and further investigation into new treatments for specific phobias is therefore important to our society, and studying factors that could affect treatment outcome is in line with this. Studying individual differences in treatment outcome is highly relevant and of interest to the individual seeking the best fitting, most economic and readily available treatment.

In summary, the results of this study show that initial impairment of phobia, and in part treatment credibility, are predictors of treatment outcome in two forms of exposure therapy for spider phobia. This confirms previous research done on predictors of treatment outcome in exposure therapy generally as well as provides novel information regarding predictors of treatment outcome in Virtual Reality exposure therapy. These results implicate that individuals with a mild to moderate spider phobia get a better treatment outcome when treated with VR OST than do individuals with a more severe and impairing phobia. Initial impairment of phobia is less important when treatment is given in form of the OST.

Acknowledgements

We would like to thank the following: our supervisor Per Carlbring, Philip Lindner for your valuable input and all involved in the Vimse research project. The VR OST application was developed by Mimerse AB, a Swedish tech-startup company that has now released the VR OST application (now named Itsy), to the consumer market (http://itsyvr.com). As Mimerse has an economical interest in the results, they were not involved in any part of the study except in the technical development and support.

Reference List

Andersson, G., Carlbring, P., & Grimlund, A. (2008). Predicting treatment outcome in internet versus face to face treatment of panic disorder. *Computers in Human Behavior*, *24*(5), 1790–1801. http://doi.org/10.1016/j.chb.2008.02.003

APA. (2013). *Diagnostic and statistical manual of mental disorders, 5th ed. (DMS-5)*. *American Journal of Psychiatry*. American Psychiatric Publishing, Inc. http://doi.org/10.1176/appi.books.9780890423349

Bailey, R. A. (2008). *Design of comparative experiments*. Cambridge University Press. ISBN 978-0-521-68357-9.

Baus, O., & Bouchard, S. (2014). Moving from Virtual Reality Exposure-Based Therapy to Augmented Reality Exposure-Based Therapy: A Review. *Frontiers in Human Neuroscience*, *8*(March), 1–15. http://doi.org/10.3389/fnhum.2014.00112

Borkovec, T. D., & Nau, S. D. (1972). Credibility of analogue therapy rationales. *Journal of Behavior Therapy and Experimental Psychiatry*, *3*(4), 257–260. http://doi.org/10.1016/0005-7916(72)90045-6

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*. http://doi.org/10.1037/0033-2909.114.1.80

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: Access, effectiveness and efficiency. Narrative literature review. *British Journal of Psychiatry*, *186*(JAN), 11–17. http://doi.org/10.1192/bjp.186.1.11

Castelloe, J. M. & O'Brien, R. G. O. (2001). Power and sample size determination for linear models A review of power concepts. *Twenty-Sixth Annual SAS Users Group International Conference*, (1992), 240–26.

Choy, Y., Fyer, A. J., & Lipsitz, J. D. (2007). Treatment of specific phobia in adults. *Clinical Psychology Review*, *27*(3), 266–286. http://doi.org/10.1016/j.cpr.2006.10.002

Coelho, C. M., Waters, A. M., Hine, T. J., & Wallis, G. (2009). The use of virtual reality in acrophobia research and treatment. *Journal of Anxiety Disorders*, *23*(5), 563–574. http://doi.org/10.1016/j.janxdis.2009.01.014

Cohen, J., Cohen, P., West, S. G., Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences* (3. ed.). Mahwah, NJ: Routledge.

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

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve Frequently Asked Questions About Growth Curve Modeling. *Journal of Cognition and Development : Official Journal of the Cognitive Development Society*, *11*(2), 121–136. http://doi.org/10.1080/15248371003699969

Davis, S., Nesbitt, K., & Nalivaiko, E. (2015). Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters. *11th Australasian Conference on Interactive Entertainment (IE 2015)*, (January), 27–30.

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in Psychology*, *6*(January), 1–9. http://doi.org/10.3389/fpsyg.2015.00026

Diemer, J., Lohkamp, N., Mühlberger, A., & Zwanzger, P. (2016). Fear and physiological arousal during a virtual height challenge—effects in patients with acrophobia and healthy controls. *Journal of Anxiety Disorders*, *37*, 30–39. http://doi.org/10.1016/j.janxdis.2015.10.007

Diemer, J., Mühlberger, A., Pauli, P., & Zwanzger, P. (2014). Virtual reality exposure in anxiety disorders: Impact on psychophysiological reactivity. *The World Journal of Biological Psychiatry: The Official Journal of the World Federation of Societies of Biological Psychiatry*, *2975*(February 2012), 1–16. http://doi.org/10.3109/15622975.2014.892632

Fredrikson, M., Annas, P., Fischer, H., & Wik, G. (1996). Gender and age differences in the prevalence of specific fears and phobias. *Behaviour Research and Therapy*, *34*(I 989), 33–39. http://doi.org/10.1016/0005-7967(95)00048-3

Hamilton, W. (2015). Spindlar höjer den verkliga hjärtfrekvensen. *Stockholms Universitet*, 1–14. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Psykologiska+institutionen#5

Hellström, K., & Öst, L.-G. (1996). Prediction of outcome in the treatment of specific phobia. A cross validation study. *Behaviour Research and Therapy*, *34*(5-6), 403–411. http://doi.org/10.1016/0005-7967(96)00004-6

Issakidis, C., & Andrews, G. (2004). Pretreatment attrition and dropout in an outpatient clinic for anxiety disorders. *Acta Psychiatrica Scandinavica*, *109*(24), 426–433.

Kaczkurkin, A. N., & Foa, E. B. (2015). Cognitive-behavioral therapy for anxiety disorders: an update on the empirical evidence. *Dialogues in Clinical Neuroscience*, *17*(3), 337–46. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4610618&tool=pmcentrez&rendertype=abstract

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593–602. http://doi.org/10.1001/archpsyc.62.6.593

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H.-U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International Journal of Methods in Psychiatric Research*, *21*(3), 169–84. http://doi.org/10.1002/mpr.1359

Klorman, R., Weerts, T. C., Hastings, J. E., Melamed, B. G., & Lang, P. J. (1974). Psychometric description of some specific-fear questionnaires. *Behavior Therapy*, *5*(3), 401–409. http://doi.org/10.1016/S0005-7894(74)80008-0

Krijn, M., Emmelkamp, P. M., Olafsson, R., & Biemond, R. (2004). Virtual reality exposure therapy of anxiety disorders: A review. *Clinical Psychology Review*, *24*(3), 259–281. http://doi.org/10.1016/j.cpr.2004.04.001

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613.

Lindner, P., Frykheden, O., Forsström, D., Andersson, E., Ljótsson, B., Hedman, E., & Carlbring, P. (2016). The Brunnsviken Brief Quality of Life Scale (BBQ): Development and Psychometric Evaluation. *Cognitive Behaviour Therapy*, *6073*(March), 1–14. http://doi.org/10.1080/16506073.2016.1143526

Magee, W. J., Eaton, W. W., Wittchen, H. U., McGonagle, K. A., & Kessler, R. C. (1996). Agoraphobia, simple phobia, and social phobia in the National Comorbidity Survey. *Archives of General Psychiatry*, *53*(2), 159–68. http://doi.org/10.1001/archpsyc.1996.01830020077009

McCann, R. A., Armstrong, C. M., Skopp, N. A., Edwards-Stewart, A., Smolenski, D. J., June, J. D., & Reger, G. M. (2014). Virtual reality exposure therapy for the treatment of anxiety disorders: An evaluation of research quality. *Journal of Anxiety Disorders*, *28*(6), 625–631. http://doi.org/10.1016/j.janxdis.2014.05.010

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. http://doi.org/10.1002/da.20734

Michaliszyn, D., Marchand, A., Bouchard, S., Martel, M.-O., & Poirier-Bisson, J. (2010). A Randomized, Controlled Clinical Trial of In Virtuo and In Vivo Exposure for Spider Phobia. Cyberpsychology, Behavior and Social Networking, 13(6), 689–695. http://doi.org/10.1089=cyber.2009.0277

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., & Altman, D. G. (2012). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *International Journal of Surgery*, *10*(1), 28–55. http://doi.org/10.1016/j.ijsu.2011.10.001

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. http://doi.org/10.1016/j.brat.2015.08.010

Muris, P., & Merckelbach, H. (1996). A comparison of two spider fear questionnaires. *Journal of Behavior Therapy and Experimental Psychiatry*, *27*(3), 241–244.

http://doi.org/10.1016/S0005-7916(96)00022-5

Nordgreen, T., Haug, T., Ost, L. G., Kvale, G., Tangen, T., Andersson, G., & Havik, O. E. (2015). Stepped care versus face-to-face cognitive behavior therapy for panic disorder and social anxiety disorder: Predictors and moderators of outcome. *Behavior Therapy*, *71*, 76–89. http://doi.org/10.1016/j.brat.2015.06.002

Ollendick, T. H., & Davis, T. E. (2013). One-session treatment for specific phobias: a review of Öst's single-session exposure with children and adolescents. *Cognitive Behaviour Therapy*, *42*(4), 275–83. http://doi.org/10.1080/16506073.2013.773062

Ollendick, T. H., & Muris, P. (2015). The Scientific Legacy of Little Hans and Little Albert: Future Directions for Research on Specific Phobias in Youth. *Journal of Clinical Child & Adolescent Psychology*, *444*(January), 1537–4416. http://doi.org/10.1080/15374416.2015.1020543

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. http://doi.org/10.1016/j.jbtep.2007.07.007

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2016). nlme: Linear and Nonlinear Mixed Effects Models. Retrieved from http://cran.r-project.org/package=nlme

Powers, M. B., & Emmelkamp, P. M. G. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, *22*(3), 561–569. http://doi.org/10.1016/j.janxdis.2007.04.006

Price, M., Anderson, P., Henrich, C. C., & Rothbaum, B. O. (2008). Greater Expectations: Using Hierarchical Linear Modeling to Examine Expectancy for Treatment Outcome as a Predictor of Treatment Response. *Behavior Therapy*, *39*(4), 398–405. http://doi.org/10.1016/j.beth.2007.12.002

R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rozental, A., Andersson, G., Boettcher, J., Ebert, D. D., Cuijpers, P., Knaevelsrud, C., & Carlbring, P. (2014). Consensus statement on defining and measuring negative effects of Internet interventions. *Internet Interventions*, *1*(1), 12–19. http://doi.org/10.1016/j.invent.2014.02.001

Scholing, A., & Emmelkamp, P. M. (1999). Prediction of treatment outcome in social phobia: a cross-validation. *Behaviour Research and Therapy*, *37*(7), 659–70. http://doi.org/10.1016/S0005-7967(98)00175-2

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. http://doi.org/10.1001/archinte.166.10.1092

Steele, F. (2014). Multilevel Modelling of Repeated Measures Data. LEMMA VLE Module 15, 1-62.

Stinson, F. S., Dawson, D. A, Patricia Chou, S., Smith, S., Goldstein, R. B., June Ruan, W., & Grant, B. F. (2007). The epidemiology of DSM-IV specific phobia in the USA: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychological Medicine*, *37*(March), 1047–1059. http://doi.org/10.1017/S0033291707000086

Szymanski, J., & O'Donohue, W. (1995). Fear of Spiders Questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, *26*(1), 31–34. http://doi.org/10.1016/0005-7916(94)00072-T

Thompson, D., Ollendick, T., & Öst, L.-G. (2012). *Intensive One-Session Treatment of Specific Phobias*. New York: Springer.

Veličković, V. M. (2012). What everyone should know about statistical correlation. *American Scientist*, *103*(1), 26-30.

Wolitzky-Taylor, K. B., Arch, J. J., Rosenfield, D., & Craske, M. G. (2012). Moderators and non-specific predictors of treatment outcome for anxiety disorders: A comparison of cognitive behavioral therapy to acceptance and commitment therapy. *Journal of Consulting and Clinical Psychology*, *80*(5), 786–799. http://doi.org/10.1037/a0029418

Wolitzky-Taylor, K. B., Horowitz, J. D., Powers, M. B., & Telch, M. J. (2008). Psychological approaches in the treatment of specific phobias: A meta-analysis. *Clinical Psychology Review*, *28*(6), 1021–1037. http://doi.org/10.1016/j.cpr.2008.02.007

Zlomke, K., & Davis, T. E. (2008). One-Session Treatment of Specific Phobias: A Detailed Description and Review of Treatment Efficacy. *Behavior Therapy*, *39*(3), 207–223. http://doi.org/10.1016/j.beth.2007.07.003

Öst, L.-G. (1987). One-session treatments for a case of multiple simple phobias. *Cognitive Behaviour Therapy*, *16*(4), 175-184. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1989-09459-001&site=ehost-live\nhttp://www.tandfonline.com/doi/abs/10.1080/16506078709455800

Öst, L.-G., Salkovskis, P. M., & Hellström, K. (1991). One-session therapist-directed exposure vs. self-exposure in the treatment of spider phobia. *Behavior Therapy*. http://doi.org/10.1016/S0005-7894(05)80374-0

Ethical considerations

This study received ethical approval from the Stockholm Regional Ethical Review Board (Dnr: 472-31). In the current study, ethical considerations can be made. According to the principles of evidence based practice, clinicians should consider and apply the best practice according to the current and available evidence (Christiansen & Lou, 2001). An argument could be made that there is not enough evidence to support the use of VR exposure therapy for individuals with spider phobia. Previous studies have shown that VR exposure therapy is effective, but many of the studies have not had ample sample sizes and a control condition. Even less research has been done on predictors of treatment outcome. Both these facts, and the fact that participants provided informed consent to participate, motivate the implementation of this study. If studies like this can provide a base of evidence for the VR exposure treatment, this would mean that more individuals suffering from spider phobia could get access to treatment.

To make sure that no participant would be at risk to worsen in their mental health condition, participants who showed indication of suicidal ideation or had a more severe/impairing complex of problems in the screening phase, were excluded from the study and followed up by the therapists involved in the study. They were also given advice on where to seek further care.

One ethical issue regarding this study is the potential for violation of integrity. This was handled by having an encrypted database where contact with participants was established as well as anonymization of participants during collection of data.

Appendix A

Treatment Credibility Scale (items rating both Virtual Reality Exposure Treatment and the traditional exposure treatment) as used in this study at pre-measurement presented below.

Dina förväntningar på **Virtual Reality** exponeringsterapin

1. Hur **logisk** tycker du att den här typen av behandling verkar?

Inte alls logiskt										Mycket logiskt
0	1	2	3	4	5	6	7	8	9	10

2. Hur **säker** är du på att den här metoden kommer att vara **framgångsrik** i behandlingen av din spindelrädsla?

Inte alls säker										Mycket säker
0	1	2	3	4	5	6	7	8	9	10

3. Med vilken grad av tillit skulle du **rekommendera** den här behandlingsmetoden till en vän med samma typ av problem som du har?

Inte alls tillitsfullt										Mycket tillitsfullt
0	1	2	3	4	5	6	7	8	9	10

4. Hur **framgångsrik** tror du att denna behandling skulle vara i behandling av andra rädslor av olika slag?

Inte alls framgångsrik										Mycket framgångsrik
0	1	2	3	4	5	6	7	8	9	10

5. Hur **förbättrad** förväntar du dig bli av den här behandlingen?

Ingen förbättring alls										Helt bra/ symptomfri
0	1	2	3	4	5	6	7	8	9	10

Dina förväntningar på den **traditionella** exponeringsterapin

1. Hur **logisk** tycker du att den här typen av behandling verkar?

Inte alls logiskt										Mycket logiskt
0	1	2	3	4	5	6	7	8	9	10

2. Hur **säker** är du på att den här metoden kommer att vara **framgångsrik** i behandlingen av din spindelrädsla?

Inte alls säker										Mycket säker
0	1	2	3	4	5	6	7	8	9	10

3. Med vilken grad av tillit skulle du **rekommendera** den här behandlingsmetoden till en vän med samma typ av problem som du har?

Inte alls tillitsfullt										Mycket tillitsfullt
0	1	2	3	4	5	6	7	8	9	10

4. Hur **framgångsrik** tror du att denna behandling skulle vara i behandling av andra rädslor av olika slag?

Inte alls framgångsrik										Mycket framgångsrik
0	1	2	3	4	5	6	7	8	9	10

5. Hur **förbättrad** förväntar du dig bli av den här behandlingen?

Ingen förbättring alls										Helt bra/ symptomfri
0	1	2	3	4	5	6	7	8	9	10