Habituation during one-session treatment- A study on within-session and within-task habituation for treatment of spider phobia

Kerstin Sindemark, Termin 10

STOCKHOLMS UNIVERSITET

PSYKOLOGISKA INSTITUTIONEN

HABITUATION DURING ONE-SESSION TREATMENT- A STUDY ON WITHIN-SESSION AND WITHIN-TASK HABITUATION FOR TREATMENT OF SPIDER PHOBIA[1]

Kerstin Sindemark

In vivo exposure has long been the gold standard when treating anxiety disorders, with its habituation and fear-reducing hierarchy model. One popular method is one-session treatment (OST) for specific phobias, which has proven to be a highly effective treatment method. However, anxiety disorders are persistent and return of fear has been observed at follow-ups despite habituating during treatment, suggesting that extinction was not achieved, calling for more effective and long lasting extinction methods. The purpose of the study was to evaluate if within-session and within-task habituation were associated with treatment outcomes in a group that received OST for spider phobia. A *linear mixed effects* model (LME) was used to model subjective units of distress (SUDS) collected at the beginning and end of each exposure trial. Changes in habituation were observed both within-task and within-session; however, these changes were not associated with treatment outcomes at post-measurement and 3-month follow-up, suggesting that habituation may not be the mechanism that causes change; instead change might occur due to other mechanisms.

Anxiety disorders are one of the most common forms of psychological disorders and are considered to be one of the largest mental health problems, primarily in the western world (Barlow, 2004; Kessler et al., 2007). They are also one of the leading causes of disability in western societies (Baxter, Scott, Vos, & Whiteford, 2013). Lifetime prevalence of anxiety disorders ranges from 4.8%- 31%, with the percentage tending to be overall higher in western countries and lower in non-western countries (Kessler et al., 2007). Craske, Stein, Eley, Milad, Holms, Rapee, and Wittchen (2017) estimate that the global 12-month prevalence for anxiety disorders is lower than the lifetime prevalence, around 14% for the span of 12 months, indicating that anxiety disorders are rather persistent over time. In a systematic review of anxiety disorders studies from 44 countries by Baxter et al. (2013) suggest that around the world about one in 14 people are clinically significantly affected by an anxiety disorder at any given time. About 18% of those that seek help via their primary care providers seek attention for anxiety disorders or for symptoms that are associated with or related to anxiety disorders (Barlow, 2004). Research shows that anxiety disorders are persistent (Craske & Stein, 2016; Pittig, Treanor, LeBeau, & Craske, 2018) with risk of return of fear after treatment (Craske & Mystkowski, 1999) and clinical significant response to treatment effects being less favorable than desired, ranging from 50-60% (Loerinc, Meuret, Twohig, Rosenfield, Bluett, & Craske, 2015; Rapee, Schniering, & Hudson, 2009). Another common problem experienced with anxiety disorders is that generalization of fear is often attributed to several different stimuli (Dymond, Dunsmoor, Vervliet,

Roche, & Hermans, 2015). Due to its prevalence and persistence, understanding the mechanisms behind anxiety disorders has been of the utmost interest for optimizing treatment (Craske et al, 2014; Craske et al, 2017; Pittig et al., 2018).

Most people have or are likely to suffer from anxiety, at some point in their lives. This can range in severity from subclinical anxiety in the form of a single panic attack or a strong sense of fear, worry in a stressful or uncomfortable situation all the way to pathological anxiety. Extensive research conducted on common symptoms that characterize different anxiety disorders include an intense or extreme fear, that can be described as "enduring" or persistent (Craske & Stein, 2016; Craske et al., 2017); avoidance of and/or experiencing anxiety during certain external or internal situations (Barlow, 2004; Craske et al., 2017); anticipating and/or interpreting external or internal situations as threatening (Barlow, 2004; Craske et al. 2017); and panic attacks (Craske & Stein, 2016; Craske et al., 2017). Many anxiety disorders are also characterized by increased heart rate, perspiration, and hyper vigilance (Barlow, 2004). Common anxiety disorders include generalized anxiety disorder (GAD), panic disorder, agoraphobia, and specific phobia (American Psychiatric Association, 2013; Craske et al., 2017). Previously, obsessive-compulsive disorder (OCD) and posttraumatic stress disorder (PTSD) were categorized as anxiety disorders in the *Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (*DSM-IV-TR), (American Psychiatric Association, 2000). Now these two disorders comprise their own separate categories in the *Diagnostic and Statistical Manual of Mental Disorders, 5th ed.* (DSM-5), (American Psychiatric Association, 2013), however, research on these diagnoses is still considered relevant due to their contributions to understanding anxiety disorders and fear acquisition (Craske et al., 2017).

Anxiety disorders are highly treatable. Despite the overall success rate of treatment outcomes, many who suffer from anxiety disorders tend to avoid seeking treatment. According to a mental health survey performed by the World Health Organization (WHO) individuals do not tend to seek treatment until about 20 years after the initial onset of an anxiety disorder, though people with GAD and panic disorder tend to seek treatment slightly earlier compared to other anxiety disorders (Wang et al., 2007). Therefore, it can be assumed that the actual percentage of people suffering from anxiety disorders is larger than reported.

Specific phobias
        Prevalence and diagnostic criteria
Specific phobias are intense fears of a certain object or of certain situations. Often, these objects or situations are avoided at all costs or endured with excessive anxiety during an encounter with the feared stimulus (American Psychiatric Association, 2013; Barlow, 2004). In DSM-5 (American Psychiatric Association, 2013) the criteria for specific phobia are listed as follows: a) the person must experience an excessive or unreasonable, persistent and/or intense fear every time they encounter the stimulus, b) the amount of fear experienced is almost immediate when confronted by the stimulus and out of proportion to the actual danger, c) the person tries to avoid situations involving the stimulus at all costs, or endures situations under extreme duress d) the fear is life-limiting and significantly impacts school, work, and/or personal life e) the fear must have persisted for at least 6 months. Clinical interviews are conducted to

determine a diagnosis. Three out of the criteria listed above must be fulfilled within the past 3 months and have persisted for at least 6 months in order to be diagnosed with specific phobia (American Psychiatric Association, 2013). In DSM-5 phobias are divided into different subcategories. They are: animal phobias such as spider or snake phobia; situational type phobias i.e. flying, driving, elevators, or being in enclosed spaces; environmental type phobias i.e. heights, deep water, and storms; blood-injection-injury i.e. seeing blood or receiving injections; and other i.e. loud sounds, choking, etc. (American Psychiatric Association, 2013; Ollendick, Raishevich, Davis, Sirbu, & Öst, 2010).

Specific phobias are considered to be one of the most common forms of anxiety disorders (Baxter et al., 2013; Wolitzky-Taylor, Horowitz, Powers, & Telch, 2008) and have the longest lifetime prevalence compared to other anxiety disorders (Barlow, 2004; Craske & Stein, 2017). In an epidemiology study on phobias and fears Agras, Sylvester, and Oliveau (1969) found that certain phobias are more prevalent in the population than others. According to a study in Barlow (2004) the lifetime prevalence for specific phobias is 11.25% whereas for other phobias, such as agoraphobia and social phobia, the lifetime prevalence is 5.63% and 2.73% respectively. A study by Kessler, Berglund, and Demler (2005) found the lifetime prevalence for specific phobias to be 12.5%, with other lifetime prevalence estimates ranging between 6-12% (Baxter et al., 2013; Craske & Stein, 2017). The age of onset for specific phobias tends to be lower compared to other anxiety disorders, ranging from between about 6-17 years of age, whereas for other anxiety disorders onset tends to be during childhood or young adulthood (Craske et al., 2017; Pittig et al., 2018). Barlow (2004) describes that the onset of phobias seems to be associated with a perceived true or false alarm to an object or situation that is likely to cause a phobic association or what Pittig et al. (2018) call *aversive associative learning* which is linked to the principles of conditioning. It is common for individuals with specific phobia to have more than one specific phobia, often causing an increase in impairment (Craske & Stein, 2017). There is a risk for co-morbidity between other psychological disorders and specific phobias, the most common being other anxiety disorders i.e. panic disorder and agoraphobia (Stinson, Dawson, Chou, Smith, Goldstein, Ruan, & Grant, 2007) but also major depression and alcoholism (Curtis, Magee, Eaton, Wittchen, & Kessler, 1998).

Avoidance behaviors are very common amongst individuals with specific phobia (Baxter, Vos, Scott, Ferrari, & Whiteford, 2014; Fredrikson, Annas, Fischer, & Wik, 1996; Rachman, 1977). *Figure 1* shows the behavioral mechanism model that leads to the maintenance of specific phobias, proposed by Öst (2013). The model describes how avoidance/escape behaviors and catastrophic thoughts lead to continued fear of a stimulus or situation. Due to the high avoidance rate, treatment is also often in turn avoided (Barlow, 2004; Wolitzky-Taylor, Horowitz, Powers, & Telch, 2008) likely causing an underreporting of the specific phobic population. A significant amount of suffering and reduced quality of life caused by avoidance behaviors is often reported amongst individuals with anxiety disorders (Baxter et al., 2014; Pittig et al., 2018), including those with specific phobia. In the case of spider phobia an individual may choose to avoid certain activities altogether such as spending time outside, going into cellars, basements, or laundry rooms, or other places where it is possible to encounter a spider, restricting the person's everyday life. Spider phobia has the highest prevalence

of all specific phobias, with an estimated lifetime prevalence of 6.1% (Öst, Stridh, & Wolf, 1998). Symptoms experienced with specific phobia are closely related to those experienced in other anxiety disorders (Craske & Stein, 2017; Craske et al., 2017). Symptoms include: rapid heartbeat, trembling, hyperventilation and/or shortness of breath (Thyer & Himle, 1987), as well as others related to fear and anxiety responses.
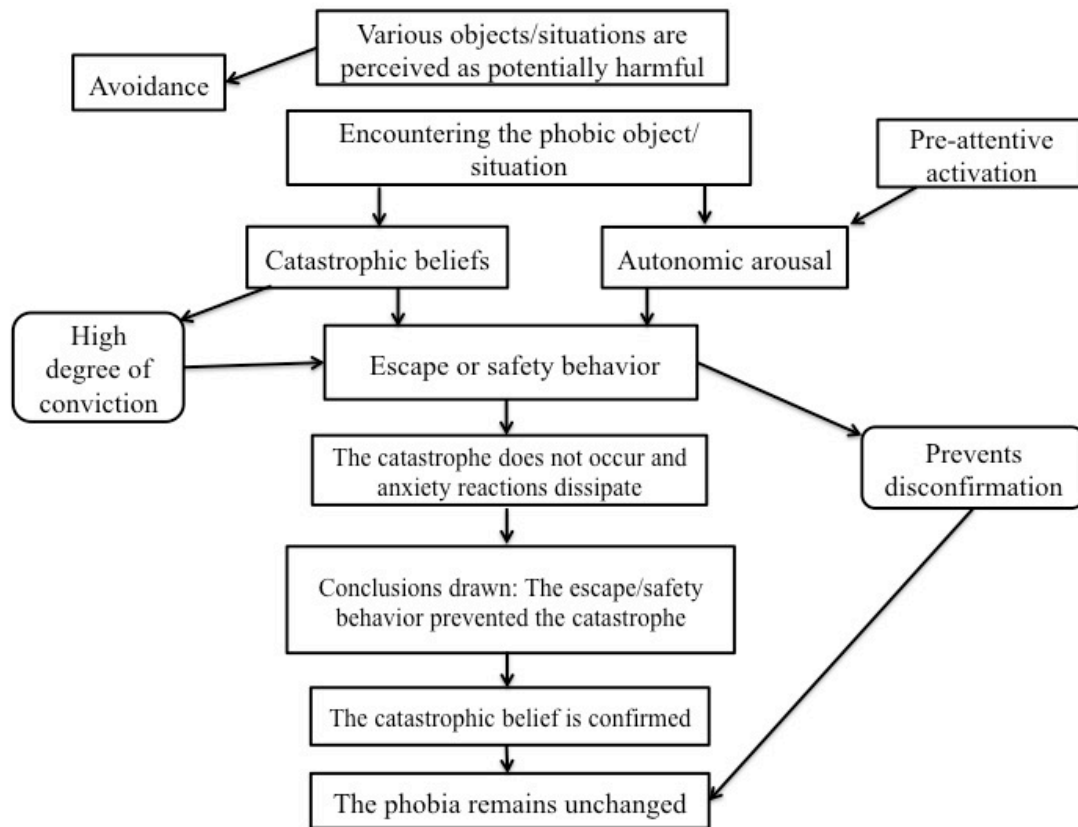

Figure 1: Lars-Göran Östs CBT model of maintenance for specific phobias. Reproduced from Öst (2013).

Etiology of respondent learning
Respondent behavioral learning theory has been one of the major building blocks of CBT and the foundation to explaining the onset and maintenance of phobias. The behavioral learning theory most used for describing the development of phobias is conditioning. There are two types of conditioning, respondent (also known as classical) conditioning and operant conditioning. Respondent conditioning occurs when an unconditioned stimulus (UCS) and a neutral stimulus (NS) are paired, causing the formerly neutral stimulus (NS) to become a conditioned stimulus (CS) and causing a conditioned response (CR). The CR is the same response that occurred for the UCS. *Figure 2* illustrates a conditioning example for acquiring spider phobia. For example a child is sitting at the table when he/she sees a black spider (NS) suddenly run across the table (UCS), and experiences an increase in heart rate, fear, and anxiety (UCR). The spider then becomes a CS and associated with the aversive outcome of feeling fear and anxiety (CR). Once this association is conditioned, later, the presence of the CS will elicit the CR. (Field, 2006; Fitzgerald, Seemann, & Maren, 2014)

Operant conditioning is related to the principles of respondent conditioning, in that operant conditioning explains how behaviors are learned and controlled through consequences, causing certain behaviors to be maintained and strengthened, and others to disappear. There are four mechanisms that explain how learned behaviors are acquired and controlled: positive and negative reinforcement, positive and negative punishment. Reinforcement increases behavior whereas punishment decreases behavior. Positive reinforcement is adding a stimulus the person likes and in that way rewarding a certain behavior, for example if a patient exposes themselves willingly to a spider or completes a difficult task in exposure they might be rewarded with a positive remark from the therapist (i.e. Good job! Well done!), or a piece of chocolate. Negative reinforcement is doing or avoiding a certain behavior or stimuli in order to avoid or stop something unpleasant, such as walking around the building instead of going through the basement door to avoid encountering a spider or turning off the TV when a spider is shown to stop unpleasant feelings like fear and anxiety. Positive punishment is adding a negative stimulus following a behavior, for example having to look at a spider for being late. Negative punishment is removing a positive stimulus following a behavior, like a child getting his/her dessert taken away for screaming when seeing a spider on the table. Ultimately, it is through these mechanisms that learned behaviors are adapted and used in everyday life. (Öst, 2013)

Respondent extinction

Just as easily as responses can be acquired they can readily fade or cease altogether. Extinction is the process of a CS being repeated with the absence of the UCS it had been previously paired with, causing the CR to reduce in frequency and decrease more and more (Fitzgerald et al., 2014; Hermans, Craske, Mineka, & Lovibond, 2006). Extinction occurs when the CS, like the spider in the example (Figure 2), is repeatedly presented to the child and nothing else fearful transpires, causing the child to become less fearful of the spider. The spider then loses its function as a CS and does not cause a CR anymore. The extinction process is considered an important component of *in vivo* exposure therapy and treatment of phobias (Craske & Myskowski, 1999; Craske et al, 2014; Foa & Kozak, 1986; Öst, 2013). However, even when extinction occurs on a previously conditioned stimulus, the conditioned response can easily be reacquired (Bouton, 2004; Field, 2006).
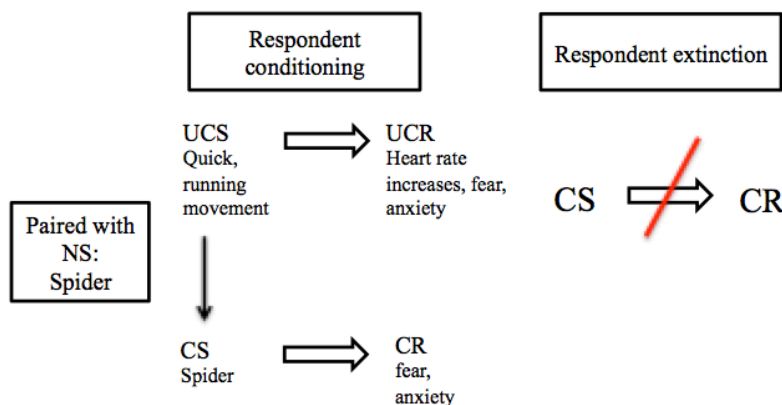

Figure 2. Illustration of how conditioning and extinction transpire.

Many find that the respondent conditioning model does not sufficiently explain the nature of phobia onset (Field, 2006; Rachman, 1977). Rachman (1977) discusses how studies have found that many phobics do not recall an initial traumatic encounter with a stimulus causing their phobia to develop, making it difficult to understand how the fear was acquired in the first place. An example of this is the case of Little Albert (Watson & Rayner, 1920) which exhibited that conditioning can occur before episodic memory has fully developed. Another factor to consider is that not every traumatic encounter with a stimulus causes a phobia to develop; casting further doubt on the conditioning model for fear acquisition (Field, 2006; Rachman, 1977). Further, Eysenck (1979) discusses how, based off of the principles of conditioning, a conditioned response to a feared stimulus (CS) should decrease if the behavior is not reinforced over time. Contrarily, phobic groups have a high avoidance rate of the CS, raising questions regarding what mechanisms actually are maintaining the fear associated with the CS. Though some have also found that phobias can develop via indirect contact, (verbal and learned information) as well as direct contact (Davey, 1992), making it possible that phobias are reinforced verbally. The above mentioned conditioning theories used to explain the conditioning of phobias were later further developed with Mowrers two-factor theory (Mowrers, 1956), which is based off of both classical and operant conditioning and further explains how phobias are maintained through avoidance behaviors. This theory has also received critique, however, in general this model seems to be widely accepted as an etiological model for phobias and fear acquisition perhaps due to its perceived simplicity and its ability to be generalized to different situations.

Clinical implications of exposure therapy
To this day the gold standard for treating phobias is *in vivo* exposure therapy (Matthews, Mackintosh, Williams, Williams, & Kirkby, 2016). Exposure therapy involves confronting a feared or anxiety provoking object or situation, often in a controlled and graded manner, to create new positive encounters with the object/stimulus and disconfirm fearful beliefs which in turn leads to new learning. One of the first to use exposure-based methods for treating phobias was Wolpe (1968) with systematic desensitization. Wolpe's (1968) method is based on the principles of conditioning and learning theory and what Wolpe (1968) calls *reciprocal inhibition*; which he describes as a relearning process that involves using or preforming a new non-anxiety producing response. The basic mechanism of the theory is that the old, aversive response is then extinguished after repeating this new non-anxiety producing response. Systematic desensitization involves gradual exposure to different phobic situations that have been identified by creating a hierarchy of phobic situations. Another component of treatment involves training in deep muscle relaxation, called progressive muscle relaxation (PMR). The PMR method is used before the patient is exposed via an imagined image of the CS in different situations and is repeated until the anxiety response has extinguished, progressing to the next, more difficult step in the hierarchy (Wolitzky-Taylor et al., 2008). In other words, Wolpe's systematic desensitization exposure method stresses the importance of habituation; i.e. to get used to the fearful situation by completely eliminating the feeling of anxiety on each step of the hierarchy before moving on to the next, which he believed was an essential component of treatment. The exposure in turn reinforces the behavior to approach feared objects. Using real anxiety-provoking stimuli, or desensitization *in vivo*, was also possible but

that method was mainly reserved for those individuals whom did not respond to the imagined situations (Wolpe, 1968).

Cognitive behavioral therapy (CBT) is one of the most common and empirically supported methods to use when treating anxiety disorders (Barlow, 2004; Craske et al., 2017; Loerinc et al., 2015; Öst, 2013) and is the primary form of therapy recommended by Socialstyrelsen in Sweden for treating anxiety disorders (2010). CBT is also one of the most effective and empirically supported therapy methods for anxiety disorders when compared to other treatment forms, (i.e. psychodynamic therapy, hypnosis, and medication), (Craske & Stein, 2016; Craske et al, 2017; Socialstyrelsen, 2010). A meta-analytic review by Tolin (2010) compared 26 different treatment studies evaluating CBT to other psychotherapy methods and concluded that at post-treatment CBT treatment in general outperformed other psychotherapy forms (psychodynamic therapies), having an overall successful outcome after treatment. Tolin (2010) found CBT to be especially superior for treating anxiety disorders and recommended that CBT should be considered the treatment of choice for anxiety disorders. Meta-analyses have found evidence to support that CBT for anxiety disorders is effective (Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016; Hofmann & Smits, 2008; Loerinc et al., 2015), even compared to waitlist (James, James, Cowdrey, Soler, & Choke, 2013). Exposure-based CBT is also the most effective intervention form for treating anxiety disorders and other disorders like OCD and PTSD (Heinig et al., 2017; Pittig et al., 2018; Öst, 2013). Treatment often occurs *in vivo*, or in real life. Generally speaking *in vivo* for specific phobias is highly effective. In a systematic review of CBT and anxiety disorders Loerinc et al. (2015) found that those with a specific phobia diagnosis achieved anxiety reduction by 52.7%. In another meta-analysis by Wolitzky-Taylor et al. (2008), the authors presented and compared several different methods to treat phobias and concluded that exposure-based treatment methods were significantly more successful compared to non-exposure based methods, and that *in vivo* may lead to quicker improvements compared to other exposure methods. Wolitzky-Taylor et al. (2008) also found evidence that supports the idea that the way exposure treatment is performed may have a significant effect on treatment outcomes.

The habituation process has since Wolpe (1968) been seen as an essential mechanism of *in vivo* exposure and that it may have significant therapeutic relevance (Barlow, 2004; Foa & Kozak, 1986; Öst, 1989). Today, *in vivo* treatment methods also follow the gradual and hierarchal approach originally used in systematic desensitization. Several other studies show that *in vivo* exposure generate successful results for treating specific phobias (Choy, Fyer, & Lipsitz, 2007; Ollendick, Öst, Reuterskiöld, Costa, Cederlund, Sirbu, Davis III, & Jarrett, 2009; Zlomke & Davis III, 2008; Öst, 1989; Öst, Brandberg, & Alm, 1997), including spider phobia, providing further evidence of *in vivo* as one of the most effective treatment methods for specific phobias.

One-session treatment
One treatment model that developed from Wolpe (1968) and systematic desensitization is the one-session treatment (OST) model of Lars-Göran Öst. OST is a manual-based treatment most commonly used for treating specific phobias (Öst, 1989; Öst, 2013) and is comprised of gradual and controlled exposure to a feared *in vivo* stimulus (Matthews et al., 2016; Zlomke & Davis III, 2008; Öst, 2010; Öst, 2013) along with participant

modeling, psychoeducation, and reinforced practice (Davis III, Ollendick, & Öst, 2009; Ollendick et al., 2009; Öst, 2013). Reinforcement is an essential component of exposure therapy since it strengthens behavior. The positive, new experiences with the phobic object reinforce the behavior to approach the object and encouragement from the therapist when performing non-phobic behavior in the session further reinforces newly acquired behaviors. The method is habituation-based, meaning the objective during treatment is to habituate or significantly decrease in fear after being exposed to the fear-evoking stimulus. It is important that a certain amount of time has elapsed and a certain amount of habituation must first have been achieved before moving on to a more difficult task (Craske et al., 2014; Matthews et al., 2016; Ollendick et al., 2009; Zlomke & Davis III, 2008; Öst, 2010; Öst, 2013). The patient must remain in the anxiety-provoking situation until the fear arousal has decreased significantly (Ollendick et al., 2009; Zlomke & Davis III, 2008; Öst, 2013). Guidelines for habituation-based exposure theory set forth by experts in the field of exposure therapy suggest the initial fear should habituate or reduce by around 50 percent before moving on to more difficult tasks in the hierarchy (Öst, 2010; Öst, 2013). This is the recommendation made and generally followed in clinical settings; however, to date, there is no data or research available that explicitly supports these recommendations.

The goal of OST is for patients to manage natural situations and display "normal" non-phobic behavior when confronted with the phobic stimuli (Öst, 1989). Non-phobic behavior is modeled during treatment by the therapist to aid in the learning process. Öst has adapted his OST treatment method to question and disprove catastrophic beliefs by confronting fears in a safe and controlled manner and ultimately, through new and positive learning experiences via performing non-phobic behavior, this then leads to the extinction of the phobia (Öst, 1989; Öst, 2013). Generally the treatment session proceeds for a maximum of 3 hours. Treatment is tailored to the client and includes a rationale for treatment and an individual-based hierarchy created of the clients' problem areas as well as some psychoeducation about anxiety and phobias. Treatment also involves the process of *overlearning*, which means to go beyond what is considered to be normal non-phobic behavior to aid in the learning process in order to enable the patient to perform the non-phobic behavior later on their own (Öst, 1989). Since reinforcement is an essential component of exposure therapy, a maintenance program is created at the end of treatment to reinforce practice after treatment and to minimize the risk of avoidance behaviors (Zlomke & Davis III, 2008). Some research has been found suggesting that maintenance programs that involve further self-exposure may to lead to maintaining positive effects after treatment (Choy et al., 2007; Hellström, Fellenius, & Öst, 1996; Hellström & Öst, 1995).

The OST method is well-established empirically and one of the most prevalent *in vivo* exposure methods used for treating specific phobias. In general, OST has about an 80-95% success rate on phobic samples (Muris, Mayer, & Merckelbach, 1998; Öst, 2010; Öst, 2013) and can be used to treat all types of specific phobias (Öst, 2013) for both adults and children (Ollendick et al. 2009; Davis III et al., 2009). Therapist-led individual OST has been found to be superior to other treatment methods such as: self-help, video treatment, and group treatment (Öst, Stridh, & Wolf, 1998), education support treatment, and waitlist (Ollendick et al., 2009). Wolitzky-Taylor et al. (2008) found in their meta-analysis that *in vivo* outperformed no treatment, placebo, and other

alternatives to exposure therapy as well as other forms of exposure. At long-term follow-ups some studies have found that the effects of *in vivo* are maintained long term about 6 months (Ollendick et al., 2009) to 1 year (Öst, Ferebee, & Furmark, 1997) and that further improvement can often occur if self-exposure continues after treatment (Choy et al., 2007). Several randomized controlled studies on treatment for spider phobia have been performed, yielding clinically significant improvement for up to 75-90% of participants (Hellström & Öst, 1995; Muris et al., 1998; Öst, 1996; Öst et al., 1997; Öst, Salkovskis, & Hellström, 1991; Öst, Svensson, Hellström, & Lindwall, 2001).

Despite the overall good treatment outcomes and the abundance of research proving *in vivo* and OST to be highly effective compared to other treatment methods for spider phobia, there are some issues with the method that lead to less than successful long-term outcomes. Arch and Craske (2009) found that a number of participants did not achieve clinically significant symptom relief after receiving exposure therapy. With other studies finding that fear can return despite having initially had successful *in vivo* treatment outcomes (Craske & Mystkowski, 1999; Ginsburg et al., 2014) causing treatment to lose its effects after a certain amount of time, and anxiety disorders to persist (Pittig et al., 2018).

Return of fear
Mystkowski, Craske, and Echiverri (2002) state that according to Rachman (1989) "return of fear is the reappearance of fear that has undergone full or partial extinction." Mystkowski et al. (2002) further explain that the return of fear can occur between exposure sessions or after a longer period of time and can be observed at follow-ups. Therefore, maintaining extinction is not certain since conditioned fear can return even after successful treatment (Craske et al., 2014). Time seems to influence the results and whether or not treatment results will be temporary or have lasting and successful results (Mystkowski et al., 2002). Craske et al. (2014) discuss that although most people do benefit from traditional exposure treatment against anxiety disorders, there are still many people whom do not benefit from treatment or whom experience return of fear after treatment. Other studies have reached similar conclusions regarding return of fear, finding it common to occur in some participants after completing exposure therapy (Craske & Mystkowski, 2006; Mystkowski, Craske, Echiverri, & Labus, 2006; Mystkowski, Mineka, Vernon, & Zinbarg, 2003). Mystkowski et al. (2002) found return of fear to occur in spider phobics when retested in a different context from where treatment occurred. Even with extensive therapy and practice after therapy studies have found that most people will not be completely symptom free and will still experience slight fear when confronted with the phobic object in their everyday life (Arch & Craske, 2009; Rachman, 1989; Rodriguez, Craske, Mineka, & Hladek, 1999). Craske et al. (2014) also found evidence that fear reduction during treatment does not predict the level of fear at follow-ups after treatment and that fear reduction is not a predictor of treatment outcomes. Craske et al. (2014) compiled research on *in vivo* exposure and discussed four different types of conditional fear which can return after treatment. These have been identified in previous studies as: fear renewal, fear reinstatement, spontaneous recovery, and rapid reacquisition.

Renewal of conditional fear occurs when the context is changed between extinction and retest (Bandarian-Balooch, Neumann, & Boschen, 2015; Bouton, 2002). If a former fearful person encounters the formerly fearful object in a context different from the context in which extinction was achieved, then that could cause a return of fear causing a new pairing and thus reactivating the conditioned response. Craske et al. (2014) draw the conclusion that this makes fear extinction seem to remain specific to the context where the extinction occurred, creating issues when treatment only occurs in a limited number of contexts. Studies have found evidence that exposure in different contexts is more effective than one context (Mystkowski et al., 2002; Mystkowski et al., 2003; Mystkowski et al., 2006) further supporting the idea that extinction is associated with treatment contexts. Reinstatement of conditional fear can occur after encountering a new aversive unconditioned stimulus between extinction and retest (Bouton, 2002). This encounter can act as a "trigger" for the extinguished fear to return. Spontaneous recovery of conditional fear can occur after not having encountered the stimulus for a certain amount of time after treatment, even after having significantly reduced in fear during treatment (Bouton, 2002). For example a person with spider phobia is successfully treated but due to the absence of spiders during the winter, they can be at risk of spontaneous recovery when the seasons change and spiders re-emerge. Rapid reacquisition occurs when re-traumatized after extinction and is more common to occur in conjunction with dangerous situations (Bouton, 2002). These complications regarding the potential return of fear and the difficulties maintaining treatment outcomes post exposure and extinction addressed by Craske et al. (2014) also raise questions of the importance and effects of habituation during *in vivo* exposure and extinction.

Habituation
Habituation-based models of exposure therapy have been a preferred method of treatment. However, the significance of habituation with regard to treatment outcomes has been much debated in recent decades (Craske et al., 2006; Craske et al., 2014; Craske, Kircanski, Zelikowsky, Mystkowski, Chowdhury, & Baker, 2008; Rowe & Craske, 1998). During exposure therapy both habituation and extinction processes are activated (Myers & Davis, 2007; Sripada & Rauch, 2015), however, in this paper habituation is operationalized as the reduction or decrease of anxiety or arousal toward a fear-evoking stimulus during exposure therapy (Matthews et al., 2016; Salkoviskis, Hackmann, Wells, Gelder, & Clark, 2006). This often occurs in conjunction with the stimulus or the situation being repeated. Habituation has been believed to be an important therapeutic response that indicates a significant decrease in fear and therapeutic change (Barlow, 2004; Foa & Kozak, 1986; Rachman, 1980), linking it to positive treatment outcomes. Most *in vivo* exposure methods are habituation-based and based on the belief that an essential mechanism for significant fear reduction is to ensure that exposure to a stimulus lasts long enough to grow accustomed to the stimulus and habituate (Craske et al. 2014). Fear-reduction (i.e. habituation) over the course of treatment sessions is also considered to be an indicator of successful treatment for habituation-based models (Benito & Walther, 2015; Davey, 1992; Foa & Kozak, 1986; Öst, 1989). Successful treatment outcomes are common in habituation-based models as previously discussed. Habituation also plays an important role in the relearning process and conditioning process (Benito & Walther, 2015). By habituating during treatment it reinforces the behavior to approach the feared stimulus in the patient, since the initial

anxiety subsides after a while, and to continue approaching it in the future. The reduction in anxiety while performing a behavior also negatively reinforces that behavior (Benito & Walther, 2015).

Within exposure therapy there are three different forms of habituation: within-task habituation, within-session habituation, and between-session habituation. Within-task habituation is defined as the amount of Subjective Units of Distress (SUDS) decreased at the beginning and end of each exposure trail. Within-session habituation is defined as the decrease in SUDS from the beginning of a treatment session to the end of the session (Sripada & Rauch, 2015). Between-session habituation is defined as the decrease in SUDS from one session to another (Sripada & Rauch, 2015), with each session often becoming easier due to the decrease in perceived fear.

A theory first proposed by Foa and Kozak (1986) called *Emotional processing theory* (EPT) explains the mechanisms behind habituation-based exposure and its effects (Foa & McNally, 1996). EPT incorporates the concept of habituation and corrective learning together create successful exposure. The theory builds off of Wolpe's (1958) theory of systematic desensitization. EPT claims that there must be an activation of a "fear structure", i.e. a combination of propositions about a stimulus, the response to the stimulus, and the association between the two. An example of a fear structure is a person seeing a spider and experiencing an increase in heart rate, and the person thinking that the spider will jump onto them. Foa and Kozak (1986) propose that during the controlled exposure, new information learned with the stimuli (i.e. spider) will create a non-fear structure. The three EPT variables that together are believed to contribute to successful exposure are: activation of fear during the exposure session, fear decreasing within the session, and fear decreasing between sessions. Together with these three EPT variables, it is believed that corrective learning can occur (Foa & Kozak, 1986). It is also believed that habituation is an indicator of therapeutic change and successful treatment (Foa & Kozak, 1986; Foa & McNally, 1996). Craske, et al. (2008) discuss in a review of EPT studies that there is very inconsistent findings regarding the EPT variables association with treatment outcomes. With some research suggesting that the variables are not functionally related to one another and do not effect treatment outcomes (Baker, Mystkowski, Culver, Yi, Mortazavi, & Craske, 2010; Peterman, Carper, & Kendall, 2016).

Further, several studies have reached the conclusion that habituation is not necessarily needed for beneficial treatment outcomes, finding that there was not a link between habituation and positive treatment outcomes (Lang & Craske, 2000; Rowe & Craske, 1998; Tsao & Craske, 2000). Instead it is believed that other factors may better explain positive treatment outcomes. Others have proposed that within-session and between-session habituation are not associated with one another (Peterman et al., 2016), which is contrary to what the EPT theory says (Foa & Kozak, 1986; Foa & McNally, 1996). Additionally, despite habituating during treatment, maintaining extinction is not certain, with risks of return of fear at long-term follow-ups causing a demand to improve extinction processes in *in vivo* exposure. Having observed these issues with *in vivo* exposure Craske et al. (2014) have proposed an alternative way of maximizing exposure through *inhibitory learning*.

Inhibitory Learning Model of Extinction

Craske et al. (2014) have compiled an extensive amount of research on habituation and *in vivo* exposure to provide insight into the function and mechanisms of exposure therapy and to draw attention to discrepancies in previous research and the need for improvement pertaining to how exposure is performed in order to create a lasting extinction process. In their review Craske et al. (2014) propose a new model based on *inhibitory learning,* which they believe to be optimal for treatment of anxiety disorders, since people with different anxiety disorders tend to lack certain inhibitory learning methods, this could lead to greater efficacy during exposure and be more effective than habituation-based approaches to exposure. Craske et al. (2014) propose that the *inhibitory learning model of extinction* can "maximize" exposure treatment and maintain more lasting results by focusing on a new extinction model. The extinction process is regarded as a central component for inhibitory learning. Contrary to the traditional idea behind extinction in *inhibitory learning*, during extinction the old conditioned association between the CS and UCS is not completely erased, instead new, or secondary, inhibitory learning develops about the connection between the CS-UCS. This gives the CS two separate meanings. (Craske et al., 2008)

Contrary to habituation-based *in vivo* exposure, where fear reduction is a key component of treatment, *inhibitory learning* does not focus on the reduction of fear in the situation but instead focuses on maintaining an elevated level of fear throughout the exposure as well as providing greater generalization to benefit the extinction process (Craske et al. 2014). Certain strategies that can be used to enhance inhibitory learning are variability in stimulus and using multiple contexts to increase generalization of the exposure therapy and counteract the risk of return of fear. Craske et al. (2014) summarize therapeutic strategies that can be used during exposure to enhance and renew inhibitory learning as: 1) expectancy violation of "frequency or intensity of aversive outcomes", 2) deepened extinction, 3) occasional reinforced extinction, 4) removal of safety signals, 5) stimulus variability, 6) retrieval cues for extinction learning, 7) multiple contexts, and 8) reconsolidation (retrieving stored memories).

Measuring anxiety and habituation

Measuring anxiety and habituation is an important part of evaluating therapy and treatment outcomes. Different methods exist for measuring anxiety and habituation. Studies have found that recording changes in physical responses such as perspiration, digit temperature, changes in heart rate, EEG scans- brain activation, changes of behavior, and self-rating scales are all successful forms of measurement (Baker et al., 2010; Hermans, Vansteenwagen, & Craske, 2006; Matthews et al. 2017; Vansteenwegen, Hermans, Vervliet, Francken, Beckers, Baeyens, & Eelen, 2005). One of the most common methods for measuring anxiety and habituation is the self-rating *subjective units of distress* (sometimes also referred to as discomfort), or SUDS (Tanner, 2012), and is a common tool used during *in vivo* exposure (Peterman et al., 2016; Öst, 2013) The scale ranges from 0-100 with 100 being the most anxiety-provoking situation the person has experienced and 0 indicating the absence of fear (Tanner, 2012; Öst, 2010). A study by Tanner (2012) found that SUDS ratings are proven to be effective in measuring the amount of anxiety experienced when encountering specific stimuli and a measure of global discomfort. The same study found that SUDS have the capability to measure both physical and emotional discomfort

(Tanner, 2012). A study found a significant relationship between SUDS-ratings, heart rate, and digit temperature (Thyer, Papsdorf, Davis, & Vallecorsa, 1984). SUDS are also simple to administer during treatment (Hermans et al., 2006) and a low-cost method compared to evaluating physiological responses. Measuring levels of anxiety during treatment is therefore a very important part of evaluating treatment methods and treatment outcomes within clinical psychology since these tools have the ability to present data and results in a manner that is directly related to the level of anxiety experienced during treatment and the habituation that occurs.

Previous research on within-session habituation and treatment results in exposure therapy
Though there is an extensive amount of research on exposure there are not all that many studies pertaining to the relationship between habituation and exposure therapy in experimental treatment settings. As of right now, most studies conducted on habituation and its relationship to treatment outcomes are on anxiety disorders in general or various single anxiety disorders (i.e. acrophobia and social phobia) and other disorders such as PTSD, with most of these studies looking at EPT variables and how they are related to treatment outcomes. Therefore, previous research is fairly limited.

In a study by Peterman et al. (2016) the authors tested how children and adolescents respond to habituation-based models of exposure for anxiety disorders. The authors evaluated the three EPT variables and found that in general the EPT variables did not predict treatment outcomes at both post-treatment and at follow-up. They did find one significant result in youth without GAD between the variable initial fear activation predicting less anxiety at follow-up. They found that between-session habituation and within-session habituation were not associated to one another. Nor were between-session habituation and initial fear activation associated with one another. They did, however find that habituation occurred during treatment but that it does not seem to be a major indicator of change.

In an experiment for social phobia using a sustained arousal exposure method, Culver et al. (2012) found that sustained arousal may be a better predictor of long-term outcomes compared to habituation of fear during treatment. They also found that variability in subjective fear during exposure might potentially promote long-term learning after treatment. They found that less within-session habituation predicted better outcomes and that outcomes relating to subjective or physiological measures of fear were not predicted by within-session habituation. The behavioral approach test (BAT) however, was associated with within-session habituation. Additionally, Culver et al. (2012) found that elevated levels of fear throughout the course of treatment and at the end of exposure, are not aversive to long-term treatment outcomes. Instead the authors propose that elevated levels of fear might actually be beneficial to improvement and enhance fear reduction long-term since their results found that elevated fear at the end of exposure predicted less fear at follow-up.

A study by Baker et al. (2010) examined habituation in an acrophobic sample, $n = 44$ using *in vivo* exposure therapy. The study examined the three assumptions or variables of EPT for maximizing exposure and its relationship to treatment outcomes. Treatment comprised of two exposure sessions with a one-week interval, a pre-test, and an

assessment directly after the last treatment, with a post-test administered two weeks after treatment. The exposure sessions varied due to predetermined time intervals linked to pre-treatment questions. Wireless heart rate monitors were used to measure pulse throughout the exposure session, SUDS-ratings were also collected at the start of the session and continued to be collected every minute. Treatment was evaluated using the BAT and SUDS-ratings. Baker et al. (2010) did not find any evidence supporting the credibility of the EPT assumptions. The results were therefore inconsistent with the EPT theory, with no obvious link found between the EPT variables and treatment outcomes. The authors further discussed how their findings raise doubts of within-session habituation acting as a mechanism of change in exposure and the limited effects that between-session habituation seems to have (Baker et al., 2010).

To the current author's knowledge, only one study to date has evaluated habituation using all the SUDS-data collected over the course of treatment. Previous studies evaluating habituation using SUDS-ratings have often used the average amount of SUDS collected during a session, or the highest SUDS-rating during treatment (peak SUDS) (e.g. Baker et al., 2010; Culver et al., 2012; Peterman et al., 2016). The study was conducted by Sripada and Rauch (2015) who, explored the effect of within-session and between-session habituation on PTSD patients receiving prolonged exposure therapy (PE) using a hierarchal linear model (LME) to analyze SUDS-data collected throughout the course of 5 minute intervals during treatment over 9 separate treatment sessions. The authors found that there was a slight increase in SUDS within treatment sessions, but that between-session SUDS decreased significantly throughout the course of treatment. Sripada and Rauch (2015) found that those who responded well to treatment, also referred to as high responders by the authors, decreased significantly more between each session, by an average of 4.37 in SUDS. The authors concluded that changes in SUDS-ratings between sessions are important for treatment and that between-session habituation benefits treatment outcomes. They also found that within-session habituation did not appear to have any significant effect relating to treatment response in their patient group and that overall, SUDS did not predict treatment outcome. Their findings did, however, also correspond to other previous research as to the importance between-session habituation has on the outcome of PTSD treatment. The authors point out that their final sample size was quite small ($n =12$) and therefore their results require replication to ensure that their treatment results are valid since the small sample size raises questions about the validity of the effects found.

Purpose
It is clear after reviewing research pertaining to exposure that there is a need for more research on habituation to examine its relationship to exposure. Craske et al. (2014) have suggested enlisting new methods to aid the extinction process to optimize exposure. However, so far there is still a need to further examine habituations' role in exposure before these methods can be applied. Some previous studies have studied the effect between-session habituation and within-session habituation has had on treatment outcomes (Baker et al., 2010; Culver et al., 2012; Peterman et al., 2016; Sripada & Rauch, 2015). Only one study was found that recently examined longitudinal SUDS-data collected over the course of treatment with the purpose to analyze the associations of between-session and within-session habituation and treatment outcomes during prolonged exposure for PTSD. Therefore, more research on habituation is needed,

especially within phobia treatment, where habituation-based exposure is commonly used (Choy et al., 2007; Wolitzky-Taylor et al., 2008; Öst, 1989) yet debates about the mechanisms within treatment that make it effective are still not completely resolved, especially regarding maintaining extinction (Craske et al., 2014). Whilst researching literature and using search words: habituation, within-session and within-task habituation, OST and spider phobia it became clear to the author that more research is needed on habituation with regard to its relationship to treatment but also that there is a tremendous need for research on the subject of within-session and within-task habituation. Only a few articles were found that explored the relationship between within-session habituation and treatment outcomes. However, no studies to date have studied within-session and within-task habituation during *in vivo* exposure therapy for spider phobia. Therefore, the importance of within-session and within-task habituation on treatment for spider phobia needs to be evaluated. A question of interest was also if treatment outcomes were maintained after treatment.

The main purpose of this paper is to study the association of habituation; which in this paper is operationalized as the reduction of SUDS-ratings; on OST *in vivo* exposure therapy outcomes for treatment of spider phobia. It also intends to examine if there is a difference in habituation between high improvers and non-high improvers, as well as to determine if improvement is stable over time and if overall treatment improvement is maintained. To the author's knowledge at the time of writing this thesis, this is the first paper to be written on the subject of within-session habituation during OST *in vivo* exposure therapy on a spider phobic sample and the first paper to evaluate if the within-task and within-session habituation changes over time.

Research questions
- Does within-task habituation change within-session?
- Does within-task and within-session habituation differ between high improvers and non-high improvers, predicting treatment outcomes?

Method

Participants and recruitment
The current study is using data that was generated during a larger experiment called Virtual Reality Method for Spider Phobia Exposure therapy or VIMSE. VIMSE is a randomized controlled trial (RCT) study that is registered in the ClinicalTrials.gov database (NCT02533310) and has received ethical approval from the Stockholm Regional Ethical Review Board (Dnr 472-31). The purpose of the larger VIMSE study was to compare traditional exposure therapy, one-session treatment (OST) for spider phobia against a virtual reality (VR) treatment for spider phobia. (Miloff, Lindner, Hamilton, Reuterskiöld, & Andersson, 2016).

Participants for the VIMSE study were recruited by various means, primarily via postings in social media and online forums, by advertisement and coverage on national television, and in newspapers where future participants found information directing them to the website: www.studie.nu and the www.iterapi.se website designed for the VIMSE project. In order to be eligible for treatment in the larger VIMSE project participants had to meet the inclusion and exclusion criteria listed on the VIMSE homepage. Participants were required to fulfill the following inclusion criteria to

participate in the study: fulfill the diagnosis criteria for specific phobia toward spiders according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (*DSM-5), live in Sweden, comprehend the Swedish language in both speech and writing, be 18 years old or older, willing to be randomized into either treatment group in the study, and have the ability to travel to the study location on three separate occasions for the pre-treatment, treatment and post-treatment (Deak & Kristoffersson, 2016; Miloff et al., 2016).

Participants were excluded from the study if they were receiving other psychological treatment at the time of the study, were pregnant, had within the last three months started taking new medication for psychological disorders, their spider fear was determined to be clinically non-significant and not meet the DSM-5 criteria for specific phobia, had suicidal thoughts, or had major depression or other problems considered too severe to receive treatment (Deak & Kristoffersson, 2016; Miloff et al., 2016). Written informed consent was obtained from all participants at the pre-measurement occasion. At the post-measurement participants were informed that those conducting the study would contact them for participation in a three-month and twelve-month follow-up of the study.

Design
The current study uses in-session collected data and selected outcome data from the VIMSE study; data collected at the pre-measurement (one week prior to treatment), treatment, post measurement (one week after treatment), and the 3-month follow-up that is relevant to the current study. The sample for the current study consists of the participants that were randomized into the OST treatment group, whereby the recruitment process and inclusion and exclusion criteria are the same as the larger VIMSE study. Table 1 presents the socio-demographic characteristics of the participants randomized to the OST treatment group that were included in the current study.

Table 1. Socio-demographic characteristics of participants randomized to OST treatment

Ost Treatment Group (*n* =45)	
Baseline Characteristics	
Age (years): *M (SD)*	33.84 (9.59)
Gender *n* (%)	
Male	8 (17.8)
Female	36 (80)
Other	1 (2,2)
Marital status *n* (%)	
Single	9 (20)
Married, registered partner	34 (75.6)
Other	2 (4.4)
Education *n* (%)	
Grade-school	2 (4.4)
High school	15 (33.3)
University or college	27 (60)
Doctoral studies	1 (2.2)
Previous psychological treatment *n* (% Yes)	12 (26.7)

Previous psychotropic treatment *n* (%Yes)	5 (11.1)
Currently on sick leave *n* (% Yes)	2 (4.4)

Figure 3 illustrates a flow chart of the different stages undergone by participants of the OST group of the larger VIMSE study. Participants answered a battery of questionnaires, self-assessment scales, and reported demographic data via the encrypted VIMSE website at registration. Once completed and the results vetted to determine whether they were allowed to move on to the next phase of the VIMSE study, participants went on to the pre-measurement. At the pre-measurement participants had to score 9 or below on the Behavioral Approach Test, also known as the BAT (see below), and meet the criteria for specific phobia set forth in DSM-5 assessed using the Structured Clinical Interview for DSM-IV Disorders (SCID-1) adapted for DSM-5 criteria; in order to be included in the study (American Psychiatric Association, 2013; Miloff et al., 2016).

The BAT was administered at three separate occasions in the study: at pre-measurement, post measurement, and at the 3-month follow-up. Participants also filled out self-rating scales during these three occasions, one of which measured fear toward spiders: Fear of Spiders Questionnaire (FSQ). After completing the pre-measurement, the larger VIMSE sample consisted of 100 participants that met all the inclusion criteria and were available for treatment. Participants were randomized into two intervention groups: traditional OST treatment or Virtual Reality (VR) treatment. Each treatment group had 50 participants assigned to them (Miloff et al., 2016). Fifty participants were originally randomized into the OST group. However, some participants were later excluded.
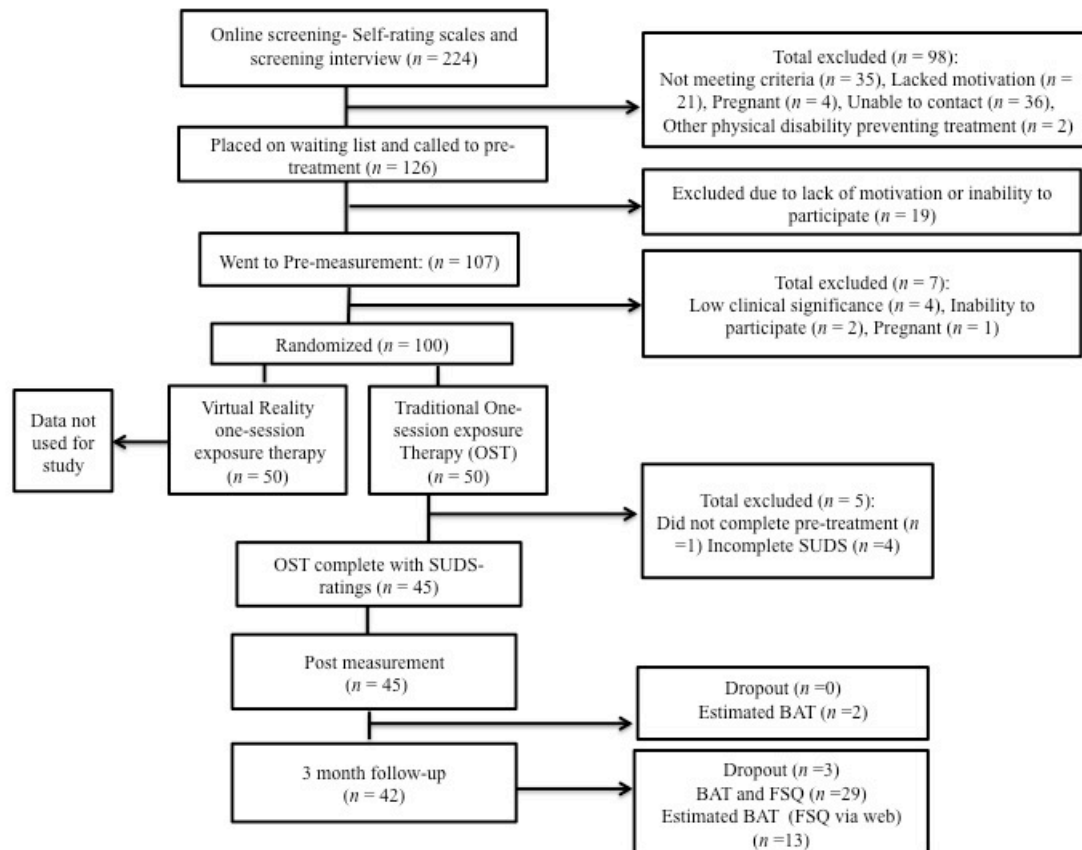
Figure 3. Flow chart of participants through the different stages of the study. FSQ=Fear of Spiders Questionnaire. BAT= Behavioral Approach Test.

Attrition

At the beginning of the study all participants (*n* = 50) completed both the BAT and FSQ questionnaire during the pre-measurement. The first dropout for this study occurred after pre-measurement and before treatment started, the reason being that the participant was no longer able to partake in the study. A total of 4 participants were excluded from the entire study due to missing SUDS-data from the OST. During the post measurement, *n* = 45 completed the FSQ questionnaire and *n* = 43 completed the BAT. During the 3-month follow-up *n* = 3 dropped out of the study, with *n* = 42 participants remaining. Linear regression was conducted to predict missing BAT values at post measurement and 3-month follow-up for participants who had FSQ scores. This method was used throughout to predict missing BAT values. No missing values were predicted if FSQ scores were missing. If neither FSQ nor BAT data was available from the follow-ups, participants were considered dropouts. For the post measurement two BAT scores were missing and predicted based off of their FSQ scores. At the 3-month follow-up 13 BAT scores were predicted using this method. Table 2 illustrates the total number of participants included in the study that completed the BAT and FSQ questionnaire at different measurement occasions.

Table 2. Total number of included participants that completed the Behavioral Approach Test and Fear of Spiders Questionnaire at different points of measurement with percentages in brackets.

Condition	Points of Measurement		
	Pre-test	Post-test	3-month Follow-up
BAT *n* (%)	45 (100%)	43 (95.5%)	29 (64%)
FSQ *n* (%)	45 (100%)	45 (100%)	42 (93%)

Procedure

Intervention: Traditional One Session Treatment

One-session treatments (OST) were conducted with clinical psychologists or clinical psychologists in their final year of training, and in accordance with the OST manual (Öst, 1989; Öst, 2010; Öst 2013). OST for spider phobia consists of *in vivo* exposure, the process of gradually exposing the participant to live or *in vivo* spiders in a planned, controlled, and graded manner, while modeling non-phobic behavior. The spiders used for the OST and the BAT, were common harmless varieties indigenous to the region of Sweden, primarily of the genus Tegenaria (T. domestica and T. atrica). The spiders were classified according to their size: small (5–15 mm), medium (15–25 mm), or large (>25 mm) and labeled size S, M, and L respectively. (Miloff et al., 2016; Öst, 2010; Öst, 2013)

OST was performed for a maximum duration of 3 hours (Öst, 1989, Öst, 2010; Öst 2013). During the session, participants were given a treatment rationale and a description of the treatment. Prior to exposure, a behavioral analysis was formed together with the participant to evaluate the participants' safety behaviors and catastrophic beliefs, and avoidance activities were also normalized. The psychologist first modeled tasks to encourage the participant to observe non-phobic behavior, as well

as to help with the relearning process. Participants were prompted to not look away or close their eyes during the modeling of non-phobic behavior and for the rest of the session. An additional half an hour was added at the end of treatment to create a maintenance plan together with the participant, making the maximum allotted time for treatment 3.5 hours.

Four different tasks were performed per spider. The first task was to teach the client how to catch the spider in a glass with the help of a postcard. The second task was for the participant to touch the spider using their finger. The third task was to let the spider walk on the participants' hand, with the goal of having the spider wander up to their elbow. Finally, the fourth task was to let the spider walk on the participants' body. All four steps are repeated for spider sizes S, M, and L, creating 12 separate tasks. However, not all the participants completed every task or made it to all three different spider sizes.

Therapists
The therapists were clinical psychologists, psychotherapists, or master's degree students in their final year of clinical training at Stockholm University, Linköping University or Karolinska Institute. Most of the therapists had already completed an OST during their clinical training, though a workshop on OST for spider phobia was held by trained clinical psychologists and by those responsible for the larger VIMSE project for all therapists before treatment began. All therapists received the same instructions for conducting *in vivo* treatment as well as a manual to follow for OST, which was reviewed during the workshop by role playing an OST session and practicing with *in vivo* spiders. Therapists were also informed that they would be able to receive additional supervision upon request. Additionally, all therapists were informed during the workshop that participants were supposed to decrease in anxiety, in this case SUDS, before moving on to a more complex task. They were also informed that they should write down participants SUDS scores before and after a task was performed and to repeat a task until the participant felt more comfortable performing it.

Outcome measure
Subjective Units of Distress SUDS is a self-rating distress scale that ranges from 0 (no distress) to 100 (maximum distress). SUDS ratings are proven to be effective in measuring the amount of anxiety the participant is experiencing when encountering specific stimuli. It can measure both physical and emotional distress, and is a common tool used during exposure treatment to evaluate anxiety (Tanner, 2012). This measure was used to evaluate habituation during the exposure treatment and to determine whether or not to proceed with the more challenging stages of treatment. (Miloff et al., 2016)

SUDS-ratings were collected over the course of treatment at the beginning and end of every exposure task, generating longitudinal SUDS-ratings. SUDS-ratings were supposed to be lower after a task was performed compared to the SUDS-ratings recorded at the beginning of the task. Tasks could be repeated several times during treatment, producing several SUDS-ratings per task. In some cases, therapists reported several sets of SUDS per task on participants; however, overall too few participants had recorded SUDS-ratings entered by the therapists on the added attempts to be able to use

the data for this study. The raw data produced by the SUDS-ratings was cleaned by the author before analyzing. The first set (beginning and end of a task) of SUDS-ratings recorded per task was used for analyzing the data. Sometimes only SUDS-ratings for the beginning of a task were recorded; in these cases the final SUDS-rating were considered *missing at random*. Contrary to instructions, several SUDS-ratings were recorded as intervals under the space provided for either SUDS at the beginning of a task or SUDS at the end of a task, e.g. 60-70. In these cases, the author consistently chose the higher number of SUDS-ratings in the interval. In certain cases, SUDS-ratings increased after a task was completed; this was a methodological error on the part of the therapist, since they did not follow the instructions to wait until the participant had habituated to record SUDS and therefore these cases were removed from the analysis. In other cases, the therapist recorded the SUDS-ratings at the beginning of a task and showed an arrow going downwards but no number, indicating a reduction in SUDS-ratings after the task was performed. For these cases only the SUDS-rating at the beginning of a task was used for the analysis. In a few cases, no first set of SUDS was recorded during the task by the therapist. However, the therapists had recorded a second set of SUDS, therefore the second set of SUDS were used for these cases. The same method was used for the cases where SUDS-ratings increased, if a second set of SUDS-ratings was available those ratings were used instead, otherwise the SUDS-ratings were left blank. Table 3 presents the percentage of SUDS recordings missing from the beginning and end of every task.

Table 3. Percentage of missing SUDS per task independent of group

	Task											
	0	1	2	3	4	5	6	7	8	9	10	11
Beginning	2%	11%	30%	41%	25%	34%	43%	61%	36%	52%	59%	73%
End	18%	25%	34%	50%	30%	41%	48%	68%	48%	57%	61%	75%

Outcome predictors
Behavioral Approach Test BAT is a standardized approach test administered in clinical and experimental settings to determine the strength of overt and observable avoidance and fear behaviors, and subjective distress amongst phobic groups (Zlomke & Davis III, 2008; Öst et al., 1991). The test is made up of 13 steps that gradually increase in grade of difficulty and was graded on a point system of 0-12, with 12 considered being non-phobic behavior. Table 4 describes in detail the different steps of the BAT. In order to be included in the study participants had to score a 9 or lower at pre-measurement, if they scored a 10 or higher their fear was considered to have a low clinical significance for treatment. In this study the BAT is used as a baseline to evaluate change in phobic behavior at pre-measurement and post measurement, and at the 3-month follow-up.

The spiders used for the BAT were medium (M) sized spiders, about 15–25 mm, and the same as those used in treatment. One size M spider was placed in a clear bin with a lid on top in the test room. Participants were informed that the BAT was not a part of treatment but instead a test to evaluate how close they could approach the spider. Before entering the room, participants were instructed that there was a spider placed in a clear bin with a lid over it. They were further instructed that the goal was to go into the room, take off the lid of the bin, pick up the spider, and hold it in their hand for at least 20 seconds. Participants were informed that they had the right to terminate the test

whenever they saw fit and were also told to try their best. Test-leaders observed from a distance to ensure that they did not in any way disturb or act as an aid to the participant. To ensure the inter-assessment validity between test-leaders; tape marks were placed on the floor to help the test-leader score the participants according to the different step specifications below.

Table 4. Behavioral Approach Test step specifications

Score	Details
0	Refusal to enter room
1	Enters room but stops before covering one-fifth of the distance to the container
2	Stops before covering two-fifths of the distance to the container
3	Stops before covering three-fifths of the distance to the container
4	Stops before covering four-fifths of the distance to the container
5	Stops before covering all of the distance to the container
6	Reaches the table with the container
7	Touches the container
8	Removes the lid of the container
9	Puts a hand inside the container
10	Touches the spider with at least one finger
11	Holds spider in hands for less than 20 seconds
12	Holds spider in hands for 20 seconds or more

Fear of Spiders Questionnaire FSQ is a self-rating scale that measures subjective fear toward spiders and uses a Likertz 7-point scale, where 1 is "absolutely not" and 7 is "absolutely". The questionnaire is proven to discriminate spider phobic groups from non-phobic groups (Szymanski & O'Donohue, 1995). The questionnaire has very good internal consistency, Cronbach's alpha = 0.92, and results are stable over about one month, according to the test-retest data (Muris & Merckelbach, 1996; Szymanksi & O'Donohue, 1995). A Swedish translated version of the questionnaire was used for the current study (Miloff et al., 2016).

Ethical Considerations
This study has received ethical approval from the Stockholm Regional Ethical Review Board (Diarienummer: 2015/472-31). The study thereby fulfills the requirements and ethical guidelines in accordance with research on humans. Participants were informed that their participation was voluntary and that they had the right to discontinue the study at any time. Participants signed a written informed consent form during the pre-measurement informing them that their personal data would be kept classified and be treated confidentially as well as handled in accordance with Swedish law. Treatment was free with the only requirement that participants had to have the ability to travel to the location of the study.

All information related to the study was handled confidentially. To ensure the confidentiality of the participants' personal data and the participants' anonymity, study codes were generated for each participant. These were generated automatically by the VIMSE website which is part of "Iterapi", a common platform used for research projects (Vlaescu, Alasjö, Miloff, Carlbring, & Andersson, 2016). The study codes were used throughout the project and were used when analyzing the data for this paper.

Participants' personal data was stored via an encrypted system on the VIMSE website. The website was protected by a two-step login in which the first code was a personal code and the second was a temporary password sent via text message. This ensured that no one outside of the study would have access to the participants' personal data as the temporary password was only sent to those who were given access to the website. No potential adverse side effects of treatment were predicted.

Analysis method
SPSS version 24 was used to calculate the data. A *linear mixed-effect model* (LME) was used to analyze how SUDS changed within-task and within-session. Linear mixed-effect models are also called hierarchical linear models or growth curve models (Field, 2018; Tasca & Gallop, 2009). LME is an effective analysis method to model longitudinal data due to its ability to handle time-varying data at both individual (random level) and group (fixed levels), as well as missing data. Tasca and Gallop (2009) describe random effects as effects that vary across units on an individual level, or across the sample. An example is the intercept representing every individual's initial baseline and the slope the rate of growth of the dependent variable. Fixed effects have a constant value for all units for all individuals and across the group (Tasca & Gallop, 2009) and presents "averages across individuals" (Hesser, 2015). LME is considered to be a better analysis method since it does not require breaking up groups into smaller groups for the sake of the analysis, as compared to ANOVAS or in the case of linear regression where different participants scores are assumed to be independent of each other (Borg & Westerlund, 2012), therefore it is a less than optimal option for modeling data with repeated measures. Instead the LME method is able to model individual change in both slope and intercept, meaning it is more sensitive and more capable of detecting change over time, and has higher power, reducing the risk for type I error. (Field, 2018; Tasca & Gallop, 2009) Maximum Likelihood was used to handle missing data in the analysis and missing data was treated under the assumption *missing at random* (MAR), (Hesser, 2015).

One of the most common methods used to measure significant change in treatment experiments is the Jacobson-Truax method for clinical significant change (Jacobson & Truax, 1991) and it was considered for analyzing the data in this paper. This method entails calculating a reliable change index (RCI) that is used as a cut-off level to determine whether clinical significant change and improvement have been achieved post treatment. This is calculated by either comparing the experimental groups post measurement results to the average score of the control group's results (Jacobson & Truax, 1991). If missing a control group or normal population to compare results to, the cut-off can be determined by using the average results of the experimental group's pre-measurements and creating a cut-off level 2 standard deviations (SD) away from the average, or M + 2SD (Hellström & Öst, 1995; Jacobson, Follette, & Revenstorf, 1984; Jacobson & Truax, 1991).

Although previous studies generated by the VIMSE project have chosen to use the Jacobson-Truax method to determine whether clinical significant change has been achieved (Dafgård, 2017; Deak & Kristoffersson, 2016) the author of this paper has opted not to use the traditional Jacobson-Truax method since almost all participants used for the current study achieved clinically significant change according to the RCI

results. Therefore, the author decided to examine the difference between high improvers and non-high improvers and to examine how many are considered to be consistent improvers over the course of post treatment and 3-month follow-up. This was achieved by creating improver variables using critical cut-off points to determine whether to classify a participant as a high improver or a non-high improver.

The BAT and FSQ questionnaire were used for determining improver variables. The improver variables were calculated by using different cutoff scores from the post measurement and three-month follow-up data (Table 5). The author chose to use a cut-off value of 34 points for the post measurement FSQ due to the experiment not having a non-phobic control group. This cut-off score was determined by calculating the median score of the participants' post treatment results, $M = 34$. Scores for FSQ were coded as 1 for high improver if they scored 34 or below and 0 for non-high improver if they scored higher than 34. For the 3-month follow-up the cutoff level for high improvers was calculated the same way, but by using the median result of the participants' 3-month follow-up, $M = 37$.

Previously, other studies evaluating BAT to determine clinical significant change have used a cutoff level of 10 points with a minimum improvement of 2 points on the test (Dafgård, 2017; Deak & Kristofferson, 2016; Öst et al., 1997). In a study by Öst et al. (1991) the maximum BAT score was used to determine if participants were considered clinically improved after treatment. In the current study, the high improver variable for the post-BAT was determined by using a cut-off value of 12, also considered to be full recovery, to ensure a more stringent method. All other results were considered to be non-high improvers. For the 3-month follow-up of the BAT for the current study participants had to score an 11 or higher in order to be considered a high improver. A decrease in one point on the BAT was allowed since the FSQ cut-off score also increased for participants who were considered to be high improvers. The decrease in the BAT score was also considered adequate since it did not surpass the frequently used cut-off of 10 points. Consistent improvers were determined by combining the post-treatment results to the 3-month results for each respective test. If participants fulfilled the criteria for high improvers on each respective test, then they were considered consistent improvers. Three participants did not partake in any of the 3-month follow up tests and therefore their consistent improver data could not be calculated.

Table 5. *Demonstrates the criteria for the different improver variables and number of participants classified in each group.*

Improver definition	Test	Cutoff	Number of High Improvers and Non-high Improvers
Improver A	Post BAT	12	High ($n = 23$) Non-high ($n = 22$)
Improver B	3-month BAT	≥ 11	High ($n = 28$) Non-high ($n = 14$)
Consistent BAT Improver	Post BAT + 3 month BAT	Must be improved both during Post BAT and 3m BAT	High ($n = 20$) Non-high ($n = 22$)
Improver C	Post FSQ	≤ 34	High ($n = 23$)

			Non-high (*n* =22)
Improver D	3-month FSQ	≤ 37	High (*n* =23)
			Non-high (*n* =19)
Consistent FSQ	Post FSQ + 3	Both of the above	High (*n* =18)
Improver	month FSQ	criteria must be fulfilled	Non-high (*n*=24)

An independent t-test was performed on the improver variables to determine whether there was a significant difference between the participants' pre-measurement scores for the BAT and FSQ and the different improver variables before further analyses were performed, see results in Table 6. A significant result was found between Improver B and the pre-BAT score of participants, $p < 0.05$. No other significant results were found.

Table 6. Results of the independent t-test

	Pre- BAT		Pre- FSQ	
Improver A	$t(43)= -0.64$	$(p= 0.52)$	$t(43)= 0.24$	$(p= 0.82)$
Improver B	**$t(40)= -2.54$**	**$(p= 0.02)$**	$t(40)= 0.84$	$(p= 0.41)$
Consistent BAT Improver	$t(40)= -1.36$	$(p= 0.18)$	$t(40)= 0.37$	$(p= 0.72)$
Improver C	$t(40)= -0.35$	$(p= 0.73)$	$t(40)= 1.29$	$(p= 0.20)$
Improver D	$t(40)= -0.35$	$(p= 0.73)$	$t(40)= 1.29$	$(p= 0.20)$
Consistent FSQ Improver	$t(40)= 0.37$	$(p= 0.71)$	$t(40)= 0.88$	$(p= 0.08)$

Results

Unconditional within-task, within-session habituation regression results

The model for the first research question was constructed using the SUDS-ratings as the dependent variable to evaluate how within-task habituation changes within-session (Table 7), regardless of improver status. Plotting the data revealed a non-linear task effect, across all three spider sizes (S, M, L), thus the time variable (task sequence) was divided into two different intervals: Time1 represents the SUDS-ratings task 0-2, while Time2 represents the SUDS-ratings between tasks 2-3. The fixed effects were Intercept, Habituation, Time1, Time2, and Spider, and interaction effects between Habituation*Time1 and Habituation*Time2. Random effects were Intercept, Habituation, Time1, and Time2.

Table 7. Unconditional and interaction effects of SUDS-ratings, estimates for SUDS-ratings according to parameter, independent of group.

Parameter	β	SE	*p*
Intercept	60.55	2.35	0.000
Habituation	-22.09	2.65	0.000
Time1	5.57	2.23	0.014
Time2	-14.76	3.36	0.000
Spider	2.41	2.27	0.291
Habituation*Time1	-3.37	1.49	0.024
Habituation*Time2	4.37	3.36	0.193

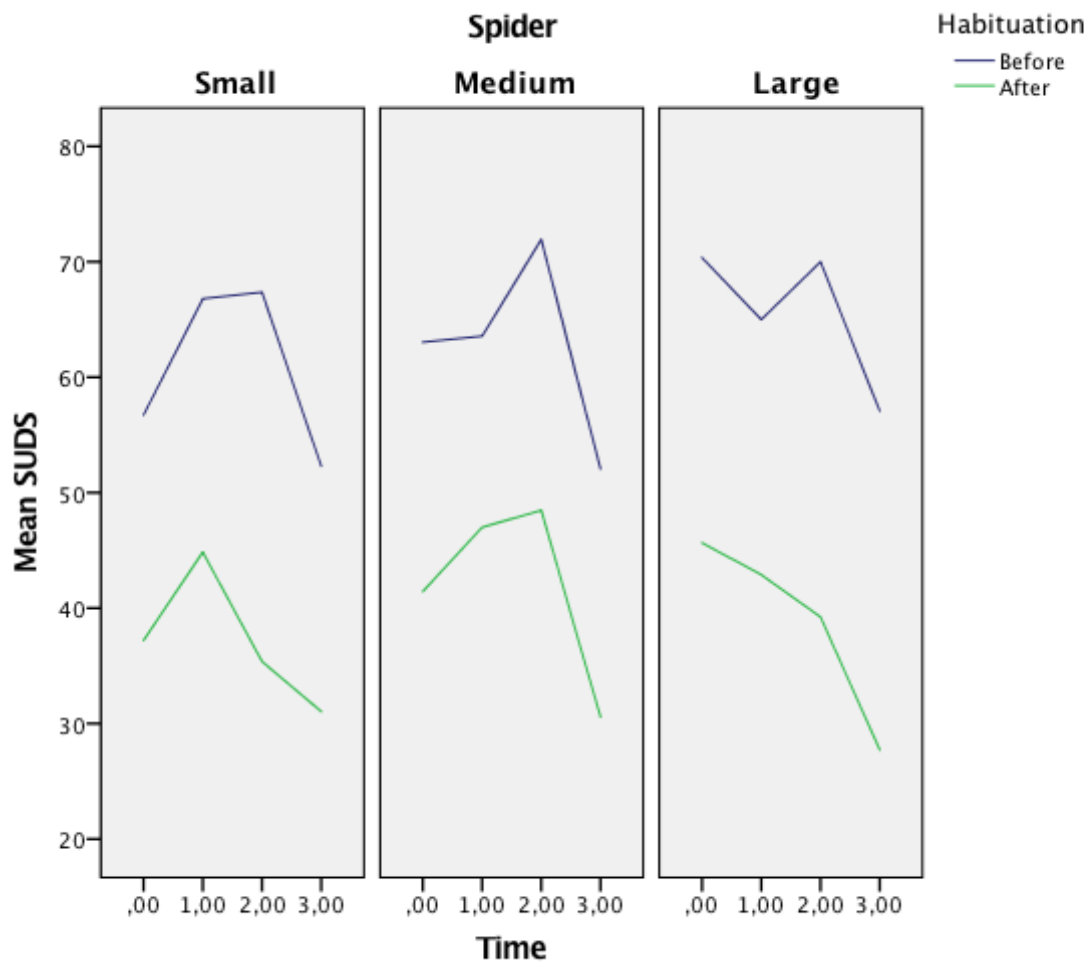*Time1= Task 0, 1, and 2. Time2= Task 2-3.*

Figure 3: Observed mean SUDS-ratings at the beginning and end of tasks, over individuals, independent of group. (Time is modeled and divided according to the four tasks per spider of OST treatment.)

The graph (Figure 3) presents the SUDS-ratings, as a function of time (task sequence) and habituation (the beginning and end of each exposure trial). The LME revealed a significant habituation effect (β= -22.09, $p < 0.001$) and that initial SUDS-ratings increased over Time1, β= 5.57, $p = 0.014$ and decreased over Time2, β= -14.76, $p < 0.001$. However, a significant interaction effect between Habituation*Time1 was found, β= -3.37, $p = 0.024$ indicating that the difference in habituation increased linearly over the first three tasks over all three spider sizes. No significant interaction effect was found between Habituation*Time2, β= 4.37, $p = 0.193$ indicating that the amount of habituation achieved from task 2-3 remained the same. At the Intercept SUDS-ratings were β= 60.55, $p < 0.001$, meaning the initial SUDS at the start of treatment for all participants were on average 60.55. There was no significant effect of spider size ($p = 0.291$).

Conditional within-task, within-session habituation
In the conditional LME models (Table 8), random and fixed effects were the same as in the unconditional analysis with SUDS-ratings as the dependent variable. The different improver variables (Improver A, B, C, D, and Consistent BAT Improver and Consistent

FSQ Improver) were added as fixed effects to the analysis, with including interaction effects Habituation*Time1, Habituation*Time2, Habituation*Improver variant *x*, Habituation*Time1*Improver variant *x*, Habituation*Time2*Improver variant *x*, the latter two denoting difference in habituation patterns dependent on improver status. Time is modeled the same as in the unconditional analysis and split up into two categories: Time1 and Time2.

The results of the LME models for the second research question show a significant effect of Habituation $p < 0.001$, indicating that the non-high improvers decreased in SUDS-ratings ranging from -20.44 to -27.74 depending on improver variable. The analysis also showed that there was no significant effect of Habituation*Improver variable effects, across the different improver variables, indicating that habituation occurred equally for high and non-high improvers. At Intercept participants who were non-high improvers started the exposure session with an initial SUDS-rating ranging from 55.50 to 67.12 depending on improver variable, with $p < 0.001$. The same pattern observed in the first analysis was observed for the results of the second analysis (Table 8) with no significant effect observed for spider and significant effects of Time1 and Time2. No significant interaction effects were found between Habituation*Time1 and Habituation*Time2 breaking the trend observed earlier in the unconditional analysis and indicating that the difference in habituation did not increase linearly in the conditional analyses. The analyses showed that no significant interaction effects were found between Habituation*Time1*Improver variable *x* and Habituation*Time2*Improver variable *x* meaning there is no significant difference in the slope of habituation during treatment between high improvers and non-high improvers (marked in gray).

Significant results were observed for Improver A (β= -12.90, $p = 0.005$) and Consistent BAT improver main effects ($\beta = -9.51$, $p = 0.047$), indicating that high improvers on the post BAT (Improver A) and Consistent BAT Improver seem to start a task with significantly lower SUDS-values compared to non-high improvers. No other significant differences in SUDS-values at the start of a task were observed for high improvers for the improver variables: Improver B, Improver C, Improver D, and Consistent FSQ Improver.

Table 8. Conditional within-task, within-session habituation, modeling the interaction effects of SUDS-ratings with Improver variables, estimates for SUDS-ratings divided according to parameter, dependent of group.

Parameter	Improver A	Improver B	Consistent BAT Improver	Improver C	Improver D	Consistent FSQ Improver
Intercept	β = 67.12, SE = 3.29 p = 0.000**	β = 62.09 SE = 4.20 p = 0.000**	β = 64.94 SE = 3.30 p = 0.000**	β = 63.69 SE = 3.35 p = 0.000**	β = 55.50 SE = 3.61 p = 0.000**	β = 58.70 SE = 3.22 p = 0.000**
Habituation	β = -25.53 SE = 3.78 p = 0.000**	β = -27.74 SE = 4.97 p = 0.000**	β = -23.90 SE = 3.77 p = 0.000**	β = -24.65 SE = 3.72 p = 0.000**	β = -20.44 SE = 4.01 p = 0.000**	β = -20.98 SE = 3.56 p = 0.000**
Time1	β = 5.52 SE = 2.22 p = 0.014*	β = 5.93 SE = 2.31 p = 0.011*	β = 5.91 SE = 2.30 p = 0.011*	β = 5.62 SE = 2.24 p = 0.013*	β = 5.87 SE = 2.31 p = 0.012*	β = 5.90 SE = 2.32 p = 0.012*
Time2	β = -14.70 SE = 3.34 p = 0.000**	β = -14.35 SE = 3.42 p = 0.000**	β = -14.36 SE = 3.41 p = 0.000**	β = -14.72 SE = 3.36 p = 0.000**	β = -14.42 SE = 3.42 p = 0.000**	β = -14.43 SE = 3.43 p = 0.000**
Spider	β = 2.41 SE = 2.25 p = 0.286	β = 2.44 SE = 2.34 p = 0.300	β = 2.40 SE = 2.33 p = 0.305	β = 2.41 SE = 2.28 p = 0.292	β = 2.50 SE = 2.33 p = 0.286	β = 2.46 SE = 2.35 p = 0.299
Improver variable *x*	β = -12.90 SE = 4.55 p = 0.005*	β = -2.30 SE = 5.07 p = 0.650	β = -9.51 SE = 4.74 p = 0.047*	β = -6.08 SE = 4.60 p = 0.188	β = 8.83 SE = 4.78 p = 0.067	β = 4.03 SE = 4.83 p = 0.405
Habituation *Time1	β = -2.55 SE = 2.23 p = 0.253	β = -3.34 SE = 2.82 p = 0.237	β = -2.94 SE = 2.17 p = 0.177	β = -3.47 SE = 2.02 p = 0.087	β = -4.09 SE = 2.17 p = 0.060	β = -3.40 SE = 1.90 p = 0.074
Habituation *Time2	β = 9.48 SE = 4.96 p = 0.057	β = 9.48 SE = 6.69 p = 0.157	β = 8.77 SE = 5.07 p = 0.084	β = 8.49 SE = 4.60 p = 0.066	β = 6.05 SE = 5.12 p = 0.238	β = 6.69 SE = 4.30 p = 0.121
Habituation*Improver variable *x*	β = 6.51 SE = 5.21 p = 0.213	β = 9.26 SE = 5.93 p = 0.120	β = 5.33 SE = 5.42 p = 0.326	β = 5.09 SE = 5.26 p = 0.334	β = -1.50 SE = 5.44 p = 0.783	β = -0.69 SE = 5.56 p = 0.901
Habituation*Time1 *Improver variable *x*	β = -1.61 SE = 2.82 p = 0.567	β = 0.11 SE = 3.23 p = 0.972	β = -0.68 SE = 2.87 p = 0.812	β = 0.13 SE = 2.79 p = 0.963	β = 1.59 SE = 2.87 p = 0.580	β = 0.37 SE = 2.93 p = 0.899
Habituation*Time2 *Improver variable *x*	β = -8.44 SE = 6.07 p = 0.165	β = -7.62 SE = 7.39 p = 0.303	β = -8.43 SE = 6.21 p = 0.176	β = -7.75 SE = 5.97 p = 0.195	β = -3.95 SE = 6.24 p = 0.526	β = -7.01 SE = 6.14 p = 0.254

* Significant interaction effect α-level: 0.05, **Significant interaction effect α-level: 0.001

Note. Improver variant *x*= Improver variant corresponding to respective improver variable

Discussion

To the author's knowledge this is the first study to evaluate whether both within-task and within-session habituation (i.e. SUDS) are associated with treatment outcomes after exposure therapy (high improver and non-high improvers). As expected, the statistical model showed a significant effect of habituation. These results are in line with previous research on *in vivo* exposure suggesting that habituation takes place during treatment (Peterman et al., 2016; Zlomke & Davis III, 2008; Öst, 1989). Habituation, however, was not associated with treatment outcomes either directly after treatment (post treatment) and later at the 3-month follow-up. These findings are more in line with Craske et al.'s (2014) research on anxiety and optimizing exposure treatment which compiles research that shows that the amount of fear reduction attained at the end of treatment, i.e. habituation, does not seem to predict the level of fear expressed later at follow-ups. It would seem, based on the results from the current study, that it is not pertinent to require a certain reduction in SUDS-ratings before moving on to the next exposure task. Further, results seem to indicate that although within-task and within-session habituation occurred, they may not be the functional mechanisms of exposure treatment that lead to successful outcomes or to the maintenance of outcomes.

 Results of analysis
Research question 1: Does within-task habituation change within-session?
A linear mixed-effect model (LME) with both random and fixed effects showed that within-task habituation changes within-session. Participants on average began tasks at a baseline of 60.55 in SUDS-ratings and decreased significantly on average by 22.09 points in habituation after a task was performed. Trends in habituation revealed that initial SUDS-ratings increased significantly over Time1, meaning that SUDS-ratings, and therefore, anxiety and fear, increased during the first three tasks per spider. SUDS-ratings later decreased significantly over Time2, between tasks 2 to 3 on all spider sizes, breaking the trend seen in Time1. However, these results conclude that initial SUDS-ratings for participants change significantly over the course of tasks and treatment.

Trends in habituation continued to be observed for all participants with a significant interaction effect between Habituation and Time1 (Habituation*Time1), (β = -3.37, *p* = 0.024) over all spider sizes, but was not observed during the interaction effect between Habituation and Time2 (Habituation*Time2). As observed in the model (Figure 3), initial SUDS-ratings continued to increase over Time1, but in general at the end of a task SUDS-ratings decreased to around 40 points indicating that habituation increased linearly over the first 3 tasks on every spider size. This further indicates that there is a significant difference in the slopes of habituation between the SUDS-ratings taken at the beginning and end (before and after) of a task on the first 3 tasks (0-2) over each spider size (i.e Time1) causing within-task habituation to change within the session.

A clear effect of habituation is observed from the results by design of the experiment, indicating that within-task habituation changes significantly within-session. Peterman et al. (2016) observed a similar trend confirming that habituation often occurs during exposure. The amount of habituation achieved in the current study within-tasks does not quite meet up to Öst's (1989) standard of habituation (50%) nor does it seem to follow Öst's (2013) recommendation for the amount of habituation that should occur during

exposure treatment before moving on to a more difficult task. This lack of amount of habituation could possibly be a result of that only the first attempt was taken into account when analyzing the SUDS-data, which is an oversight of the author. It is possible that if all trial attempts were taken into account the amount of habituation would have been near 50%. The results from this study instead suggest that participants remained fairly stable over the course of treatment with an average decrease in habituation by 22.09, which is not a decrease of 50% from the initial SUDS-ratings ($\beta = 60.55$). These results are more in line with Craske et al.'s (2014) recommendations for exposure through *inhibitory learning model of extinction*. In their paper on optimizing exposure, Craske et al. (2014) state that fear reduction is not the goal of the method instead a maintained elevated level of fear is desired throughout the extinction process further explaining that arousal and fear should therefore remain fairly stable over the course of exposure and not increase or decrease too much during exposure.

Research question 2: Does within-task and within-session habituation differ between high improvers and non-high improvers, predicting treatment outcomes?

Separate LME models were constructed for the respective improver variables. The analyses did not demonstrate any significant interaction effects between Habituation, Time1 or the different improver variables, with no significant interaction effects found between Habituation, Time2 and the improver variables (marked in grey in Table 8). Hence, the results demonstrate that within-task and within-session habituation does not seem to differ between high improvers and non-high improvers, nor are within-task and within-session habituation associated with treatment outcomes. These results can be compared to those found in the study on anxiety in youth and acrophobics which concluded that habituation was not a major predictor of change despite habituation occurring during exposure (Baker et al., 2010; Peterman et al., 2016). Similar results were found in Culver et al.'s (2012) study on public speaking that indicated outcomes relating to subjective measures of fear were not predicted by within-session habituation; which can be compared to the results of Improver variables C, D, and Consistent FSQ Improver which were based off of results from FSQ questionnaire. However, the results of the current study stride against those found in Culver et al. (2012) regarding within-session habituation being a predictor of BAT outcomes, with no significant interaction effects observed for Improver variable A, B, and Consistent BAT improver with Time1 and Time2, and Habituation indicating no association between habituation and BAT treatment outcomes. Differences in methods and participants could however, have to do with the contradictory results.

Sripada and Rauch's (2015) study did not find strong support for an association between within-session habituation and positive treatment outcomes. These results are similar to those presented in the current study, with the current study finding no association between within-session and within-task habituation and treatment outcomes. However, Sripada and Rauch's study has several noteworthy limitations. Their sample size was small and their results may have been affected by the amount of power generated by the low number of participants (*n*=12). The sample size also limits the amount of SUDS-ratings generated in correspondence to treatment. Perhaps if Sripada and Rauch (2015) had had a larger sample size, they might have been able to reach more conclusive results between the effect of within-session habituation and positive

treatment outcomes. Therefore, it is difficult to draw any conclusive results on the nature of within-session habituation from Sripada and Rauch's (2015) study in relation to the current study.

High improvers for the different Improver variables *x* started a task either with lower or higher SUDS compared to non-high improvers, but with no significant differences observed between the amount of habituation, with the exception of two improver variables. The analyses for improver variable A (post BAT) and Consistent BAT improver yielded significant results with $\beta = -12.90$, $p = 0.005$ and $\beta = -9.51$, $p = 0.047$ respectively. These results indicate that high improvers for Improver Variable A and Consistent BAT Improver started a task with SUDS-ratings significantly lower compared to non-high improvers. Similar results to the unconditional analysis were observed with no significant effect of spider observed, with significant results observed over all improver variables for the parameters: Intercept, Habituation, Time1 and Time2, with $p < 0.05$.

Further reflection on the findings
Overall, treatment was successful for all participants. This result was in accordance with expectations and with previous studies that have found conclusive evidence of *in vivo* exposure's effectiveness on specific phobias such as spider phobia (Choy et al., 2007; Ollendick et al. 2009; Zlomke & Davis III, 2008; Öst, 1989) providing further evidence that OST for phobias is the "gold standard" within exposure therapy. However, contrary to previous research (Deak & Kristofferson, 2016; Dafgård, 2017; Hellström & Öst, 1995; Öst 1997; Öst et al., 1991) clinical significant change was not used to evaluate the current study's results. This was due to the fact that the majority of participants in the current study significantly improved and obtained clinical significant change. Thus, it called for a new method for evaluating improvement amongst the participants creating the distinction between high improvers and non-high improvers. Since the criteria for high improvers was more stringent than that for achieving clinical significant change, it could be argued that results from the current study can be compared to previous studies that used clinical significant change to evaluate treatment outcomes. Further, the results of the study indicate that improvement remained fairly stable over time and that most participants who improved continued to consistently improve at follow-ups. These results follow trends with previous research claiming that participants show improvement at follow-ups after receiving *in vivo* exposure (Choy et al., 2007; Hellström et al., 1996; Hellström & Öst, 1995).

The tendency to habituate within-task and within-session was observed for all participants in the current study. This tendency was already observed in treatment due to the experiment's design. However, some significant differences in habituation were observed within-task and within-session between high improvers and non-high improvers. With non-high improvers having a tendency to significantly habituate and reduce in SUDS between 20.44-27.74 points after a task was performed, depending on Improver variable. Non-high improvers also tended to increase significantly in SUDS-ratings at the beginning of the first three tasks (Time1) between 5.52-5.93 points. This trend was broken over Time2, where non-high improvers tended to decrease significantly in SUDS-ratings between 14-35-14.72 points between task 2-3 compared to high improvers.

It is important to note that the results of the current study indicate that although there sometimes was a significant difference in habituation between high improvers and non-high improvers, this difference was not related to treatment outcomes. Thus, leading to the conclusion that within-session and within-task habituation may not be the mechanisms that cause exposure to be successful. Instead, rather arguing for other possible mechanisms of change, such as how treatment is performed (Wolitzky-Taylor et al., 2008) and *inhibitory learning model of extinction* (Craske et al., 2014). No speculations regarding whether the amount of habituation observed in the current study could be considered by Craske et al. (2014) as an elevated level of arousal since it was not specified what would be considered elevated levels of anxiety over the course of exposure.

Further, the effect of the maintenance plan implemented at the end of treatment could possibly be one factor that separates high improvers from non-high improvers. Practicing and self-exposing oneself after *in vivo* exposure has been found to help maintain the positive effects attained after treatment (Choy et al., 2007; Hellström et al., 1996; Öst et al., 1991). No data pertaining to this factor was collected over the course of the study making it impossible to assess if there was a difference between the two groups.

Methodological considerations
    Outcome measures
This paper is based on already collected data from the larger VIMSE study and therefore the possible outcome measures for the experiment were limited to the data collected. SUDS-ratings were used as the outcome measure for habituation possibly leading to issues with external validity for the study. SUDS-ratings are subjective and therefore participants' previous life experiences effect what they would consider to be their most fearful experience when encountering a spider with their SUDS-rating adjusting accordingly. Another potential concern for the external validity of the study is that participants can sometimes conform to what is considered social desirable in therapy settings when asked leading questions (Fisher, 1993). Therefore it is possible that participants reported a decrease in SUDS to attempt to please the therapists and do what is expected of them. However, SUDS-ratings have been used in several studies to evaluate habituation (Baker et al., 2010; Culver et al., 2012; Peterman et al., 2016) and have been found to be sensitive to changes in fear and anxiety experienced during *in vivo* exposure (Tanner, 2012). Questions regarding the reliability of therapists documenting SUDS-ratings during treatment arose in the current study. Although all therapists had received the instruction to record SUDS-ratings at the beginning and end of every task, this instruction was not followed for some of the participants. A total of four participants did not have any SUDS-ratings recorded, causing the participants to be excluded from the study. Other therapists had only recorded SUDS-ratings at the beginning of a task, whereas in some cases therapists recorded SUDS-ratings increasing after a task was performed. This last issue was an error on the part of the therapist since SUDS-ratings are supposed to decrease or at least remain stable during exposure, not increase, indicating that the therapist asked too early for the SUDS-rating and did not wait until the end of the task to ask for the SUDS. Another possible reason for the discrepancy in SUDS-ratings may be that documenting the SUDS-ratings may not have

been a high priority amongst the therapists and that the therapists forgot, neglected, or grew tired of collecting SUDS-ratings during treatment leading to the possible threat of the internal validity.

The therapists in the current study received instructions to follow Öst's (1989) recommendations to habituate 50% before moving on to a more difficult task. However, upon examining the raw data it seems like this recommendation was not generally observed by the therapists. The therapists did however, wait for a significant decrease in SUDS before moving on to the next task, showing that habituation still occurred. It is also possible that the SUDS-ratings could differ depending on the therapists, causing a therapeutic effect, since it ultimately is up to each individual therapist to decide when the participant is ready to proceed to the next stage in treatment. Therefore, habituation might possibly vary depending which therapist performed the OST. Potential therapeutic effects were not evaluated in this study. The amount of SUDS collected during treatment decreased considerably over the last two tasks for spider size L as indicated by Table 3. The consequence is that it leads to uncertainty as to what occurs with the trends in habituation with regard to the last spider (L) on the last two tasks.

The variables used for the improver variables were chosen due to their proven ability to evaluate changes in phobic behavior and all measures were used in the pre-treatment, post treatment and the 3-month follow-up. BAT and FSQ are believed to be complimentary to each other, with BAT providing information of observed phobic behavior and how physically close a participant can get to the phobic object, whereas FSQ provides information of the participant's subjective fear towards spiders. FSQ has good psychometric properties (Muris & Merckelbach, 1996; Szymanski & O'Donohue, 1995) and is able to differentiate spider phobics from non-spider phobics (Szymanski & O'Donohue, 1995). The use of BAT as an indicator of improvement raises questions of the test's face validity. It has become standard to use BAT when following-up treatment studies for specific phobia (Hellström & Öst, 1995; Muris et al., 1998; Öst, 1997; Öst et al., 1991; Öst et al., 1997; Öst et al., 2001). However, there is no research pertaining to the psychometric properties of BAT, making it difficult to conclude if the BAT actually measures what the test claims. Inter-rater reliability between therapists rating the BAT should not be a methodological issue for the current study since tape markers were placed according to distances used by Öst (1991) previously for the BAT.

During treatment, but also occasionally during the BAT the spiders could die. This could happen due to them getting squished or from frequently being used or due to other reasons. The therapists were told to have several spiders on hand in the different sizes in case spiders died during the course of treatment. This was taken as a precaution to ensure that the internal reliability remained high. There is only one recorded instance of treatment not being able to go on due to the lack of spiders in the correct size, but it is possible that it happened other times thus potentially threatening the internal reliability of the treatment.

Internal validity
A possible threat to the internal validity is that participants had to physically show up for the post treatment and 3-month follow-up of the BAT. Those who did not show up to the follow-ups received an online version. The BAT could not be administered online

due to the nature of the test. Linear regression was used to estimate the missing BAT scores. This is an effective method to use when independent data is missing (Field, 2018; Pallant, 2013) however, raw BAT data is naturally preferred. No estimated score was designated a 12 for either the post treatment or 3-month follow-up. Some BAT scores for the 3-month follow-up were estimated to 11, making these cases meet improver criteria at the 3-month follow-up. This should not affect the statistical analysis considering that it pertained to only a few cases and the estimated values were the same or one point less than what participants had scored during the post treatment.

Another threat to the internal validity of the study is that therapists received the instruction that participants should habituate by about 50% before moving on to the next step in treatment. This instruction could have had an effect on the SUDS-ratings if the therapists had participants remain longer in a situation, possibly creating a larger gap in habituation obtained upon completing a task than it would normally. This does not seem to be the case upon examining the raw data, however, since the majority of the SUDS data did not decrease by 50% during tasks. The author recognizes that this instruction may not have been optimal for the current study's method and that this is something to consider for future studies. A possible indicator of therapists having participants remain on the same task longer is the number of tasks completed throughout the OST and the number of times a task was repeated.

External validity
A threat to the external validity of the current study is the absence of a control group with which to compare FSQ and BAT results, and therefore, the ability to generalize the results. It may be helpful to the external validity of the study to compare the cutoff-levels of FSQ and the BAT. Due to the absence of the control group the cutoff values the FSQ for the current study were created based on the participants' average score from post treatment instead of calculated according to clinical significant change. The cut-off value for improvement on the BAT has in previous studies been 10 (Dafgård, 2017; Deak & Kristofferson, 2016; Öst et al., 1997). The current study used higher scores at the post-test and 3-month follow-up, 12 and 11 points, respectively. The more stringent score was chosen due to the unusually high number of participants that received 10 points or higher on the BAT.

Another possible threat to the external validity of the current study is that participants volunteered for the study, rather than being remitted. Participants however, were randomized from the larger VIMSE study to receive OST, therefore the fact that they volunteered for treatment should not affect the external validity. As mentioned earlier, individuals with anxiety disorders have a tendency to avoid treatment, raising questions regarding if the sample group is representative of those with spider phobia. Also, in order to receive treatment participants had to be able to travel to a university in the Stockholm area, possibly limiting the sample to being only representative of those living close to the area and not of all of Sweden. However, the clinical interviews performed during the pre-treatment ensured that the study sample fulfilled a specific phobia diagnosis for spiders and the study should therefore provide, in general, a representative picture of those with spider phobia.

Aspects of the statistical analysis

A *linear mixed-effect model* (LME) was the statistical method chosen to analyze the data due to the method's advantage of handling missing data and the capability to model longitudinal data. Various other options to model and analyze the data exist, e.g. linear regression. However, a disadvantage with linear regression is that different participants scores are assumed to be independent of each other. This would have been problematic for the current study seeing as how the study aimed to model repeated measurements of the same data (SUDS) and therefore linear regression would not have been a viable option to present the complete relationship between SUDS-ratings to all of the different variables. It was therefore decided that the LME was a more suitable option due to the benefits of the model, such as not needing to assume homogeneity of regression slopes and its ability to handle missing data (Field, 2018) and would model the SUDS-data well. Although tasks could be performed several times throughout the study, generating more SUDS-ratings per task, the number of missing SUDS on second attempts largely exceeded the number of SUDS collected on the first attempt. Therefore, only the first attempt (beginning and end) of a task was used in this study. It would also have been considered a methodological error to use the second attempts on tasks, since the individuals who wanted to try a task again presumably did not feel that they habituated during their first attempt, possibly adding bias to the analysis since the SUDS-data would no longer be considered *missing at random* (MAR) for the participants that only made one attempt at a task.

A factor that could possibly have affected the results of the current study is that the improver variable groups might have had too few participants in some of the high improver and non-high improver groups. Field (2018) recommends that when using LME analyses, group sizes should not be considered to be too small. This is a fairly vague recommendation and recommendations for actual group sizes were not specified; however, it leads to the possibility that the results regarding the improver variables in the current study may have been different if high improvers and non-high improvers had been more evenly distributed or if the study had had more participants. The LME handled a total of 1056 SUDS-ratings for the entire study with 606 SUDS-ratings generated during treatment and 450 SUDS-ratings estimated by the LME model compared to Sripada and Rauch's (2015) study which generated 339 SUDS throughout the course of treatment. The current study presumably generated far more power from the sample (post: $n = 45$ and 3-month follow-up: $n = 42$) on both random and fixed levels than Sripada and Rauch's study, considering the recommendations for performing LME analyses is to not have too small group sizes and Sripada and Rauch's sample size was considerably small ($n = 12$). One strength of the current study however; is that the study included several measurement occasions for the outcome measurement (SUDS) during the intervention phase, with a total of 24 measurement occasions possible (at the beginning and end of a task). In LME analyses it is beneficial to have several measurement occasions during the intervention phase especially when examining mediators and moderators (Hesser, 2015).

Study limitations
This is the first study of its kind and therefore, results should be interpreted with study limitations in mind. One study limitation is the absence of a control group. This is due to the experimental design of the larger VIMSE project. However, Sripada and Rauch's study (2015) also lacked a control group from which to evaluate potential outcome

differences. In the future, studies evaluating longitudinal habituation data could consider a control group where a certain amount of habituation is to occur before moving on to the next phase of treatment and test against an experimental group where no instructions on the amount of habituation is given. Another possible limitation to the study is the sample size in relation to the improver variables. The group size of the current study, could not be amplified due to the study design of the larger VIMSE study limiting the overall sample size.

Several limits are due to the current study having used data already collected and compiled during the larger VIMSE study, such as the materials used for the evaluating treatment outcomes and SUDS-ratings. In the larger VIMSE study SUDS were the only measure used to record habituation. Therefore, no data pertaining to physiological responses to fear was used in the current study. Hermans et al. (2006) discuss in their article on fear and learning that using self-report instruments to measure fear is somewhat limiting to fear research. Stating that fear and fear learning involve both behavioral and physiological measures that should be measured with appropriate instruments (Hermans et al, 2006). This could therefore be considered a limit to the current study seeing as how no physiological measurements were taken. This should be taken into consideration for future studies. However, for the current study SUDS should be considered sufficient due to their ability to measure changes in fear and anxiety (Tanner, 2012).

Another limit is that only post measurement data and 3-month follow-up data were used to evaluate habituation improvement over time. Data from the 12-month follow-up of the VIMSE study existed but was not used, due to further attrition of participants, ($n =$ 40), making the group sizes even smaller for evaluating the improver variables and possibly further affecting the reliability of the results. It is not uncommon for dropout rates to increase at follow-ups making it difficult to interpret longitudinal data for studies. Future studies should consider using data collected over a longer period of time to evaluate if trends in habituation continue over a longer span of time. Another consideration might be to include some type of incentive to encourage participants to travel for follow-ups, so that data that cannot be evaluated online, as in the case for BAT data, can be collected rather than estimated.

Also, several other factors could potentially have affected the results of the study, but were not evaluated in the current study. These are: the length of the session, therapist contact, the maintenance program and the number of tasks performed per spider. Treatment was limited to a maximum of 3 hours; however, treatment could have been shorter from participant to participant. The study did not assess if treatment length of the session was associated with high-improvers and non-high improvers. Therapist contact can have different therapeutic effects. Inter-rater reliability was not performed on the therapist performance of OST. Studies have shown that discrepancies can be found between therapists in studies causing a difference in improvement amongst participants (Barlow, 2004). The discrepancy of inter-rater reliability between therapists should not be all that large considering all therapists completed the same treatment workshop and had previous experience performing OST. What possible effect the maintenance plan may have had on outcomes was not evaluated. Participants formulated a personal maintenance plan with their therapist prior to completing

treatment. Whether or not participants followed their maintenance plan and how often they continued exposing themselves after treatment was not taken into consideration in the current study. However, it is well-known that repetition of behaviors is good for maintaining them and that maintenance plans help maintain newly learned behaviors (Öst, 1989) and may contribute to continued improvement in individuals after treatment (Choy et al., 2007; Hellström et al., 1996; Hellström & Öst, 1995; Wolitzky-Taylor et al., 2008). Therefore, the maintenance program may have been interesting to evaluate. Not all participants completed all of the 12 possible tasks during treatment, some may only have completed certain tasks and continued practicing the remainder of the exposure making it difficult to determine what actually occurred on the last tasks pertaining to the last spider (L). However, exposure therapy is customized to the individual and what they feel comfortable doing so it is not surprising that not all participants completed all of the steps of the treatment since treatment is set up this way.

Clinical implications and future research
The clinical implications that can be drawn from the study at hand should be done carefully due to the aforementioned limitations of the study and with respect to this being among one of the first studies to evaluate the effects of within-task and within-session habituation on spider phobia. The results suggest that habituation was not associated with treatment outcomes, leading to the possibility that habituation may not be the mechanism that determines treatment outcomes and that they, further, may not be an indicator of whether treatment has a lasting effect. This is in line with previous research compiled by Craske et al. (2014) which questions the habituation-based approach to exposure therapy, concluding that there does not seem to be an association between the amount of fear decreased during treatment and lasting treatment outcomes. However, more research is needed to determine what the functional mechanisms of exposure are that lead to lasting change and improvement.

Research has found that anxiety disorders are amongst the most debilitating psychological disorders with high life-time prevalence rates, causing significant suffering and often, high avoidance of anxiety provoking situations, making it pertinent for the patient group to receive suitable and effective treatment. Research has also suggested that there is a need to optimize the extinction process of treatment for anxiety disorders, since it has been noted that many upon long-term follow-ups experience return of fear or a number of participants are no longer considered to be clinically significantly improved. Hence, further research is required on all anxiety disorders in order to optimize extinction processes in treatment and to accumulate more knowledge of the functional mechanisms behind exposure therapy. With more conclusive research, it could perhaps one day lead to a change in how exposure therapy is performed within the field of CBT, with clinical professionals moving away from hierarchal and habituation-based exposures to new methods of exposure that produce more generalization during extinction such as the *inhibitory learning model of extinction* proposed by Craske et al. (2014).

Furthermore, the study raises questions regarding the ability to duplicate results in future studies and what type of results studies testing different phobias might generate. A potential future research question might be to evaluate the results of two treatment

groups that received the same intervention (OST), whereby the experimental group is supposed to remain at a constant level of fear throughout the exposure and the control group is to habituate (beginning and end of a task). Measuring participants pulse responses to fear as well as collecting SUDS during treatment could provide more reliable data and minimize the risk for participant's conforming to the social desirability of reducing in SUDS.

Conclusion

The findings from this study provide preliminary evidence suggesting that the habituation process experienced during OST for spider phobia, is not associated with treatment outcomes (high improvers and non-high improvers). The results also indicate that habituation was observed over the course of treatment with all participants significantly decreasing in fear while performing treatment tasks, providing evidence that habituation occurred. The results also found evidence that supports the idea that within-task habituation changes within-session. Habituation was also observed for both high improvers and non-high improvers. Further, the results indicate that improvement (high improvers and non-high improvers), both at post treatment and at the 3-month follow-up, does not seem to be related to the amount of habituation experienced during treatment. This is in line with research suggesting that habituation and the reduction of fear during treatment may not be an indicator of treatment outcomes (Baker et al., 2010; Craske et al., 2014; Culver et al., 2012; Peterman et al., 2016), but instead, may be due to other mechanisms in the treatment process or how treatment is performed (Craske et al., 2014; Wolitzky-Taylor et al., 2008). What these potential mechanisms may be falls outside the scope of this study. The study confirms that OST is a highly effective intervention to use when treating spider phobia, exhibiting exemplary results and providing further support of the method's superiority and "gold standard" classification. However, the study results also provide evidence that support the idea that the habituation process may not be the working mechanism of *in vivo* exposure, which strides against the concept behind habituation-based exposure models, but instead finding results that coincide with Craske et al. (2014) and their recommendations for maximizing exposure using the *inhibitory learning model of extinction*.

References

Arch J. J, & Craske M. G. (2009). First-line treatment: a critical appraisal of cognitive behavioral therapy developments and alternatives. Psychiatric Clinics of North America. 32(3):525–547.10.1016/j.psc.2009.05.001 [PubMed: 19716989

Agras, S., Sylvester, D., & Oliveau, D. (1969). The epidemiology of common fears and phobia. Comprehensive Psychiatry, 10(2), 151–156. http://doi.org/10.1016/0010-440X(69)90022-4

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

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

Barlow, D. (Ed.). (2004). *ANXIETY and Its Disorders* (Second Edition). New York: Guilford.

Bandarian-Balooch, S., Neumann, D. L., & Boschen, M. J. (2015). Journal of Behavior Therapy and Exposure treatment in multiple contexts attenuates return of fear via renewal in high spider fearful individuals. *Journal of Behavior Therapy and Experimental Psychiatry*, *47*, 138–144. http://doi.org/10.1016/j.jbtep.2014.12.006

Baxter, A. J., Scott, K. M., Vos, T., Whiteford, H. A., (2013). Global prevalence of anxiety disorders: a systematic review and meta-regression. Psychol. Med. 43, 897–910. http://dx.doi.org/10.1017/S003329171200147X.

Baxter, A. J., Vos, T., Scott, K. M., Ferrari, A. J., & Whiteford, H. A., (2014). The global burden of anxiety disorders in 2010. Psychol. Med. 44, 2363–2374. http://dx.doi.org/10.1017/S0033291713003243.

Benito, K. G., & Walther, M. (2015). Journal of Obsessive-Compulsive and Related Disorders Therapeutic process during exposure : Habituation model. *Journal of Obsessive-Compulsive and Related Disorders*, *6*, 147–157. http://doi.org/10.1016/j.jocrd.2015.01.006

Borg, E., & Westerlund, J. (2012). *Statistik för beteendevetare. Faktabok* (3e uppl.). Stockholm: Liber.

Bouton, M. E. (2002). Context , Ambiguity , and Unlearning : Sources of Relapse after Behavioral Extinction Four Mechanisms of Relapse, *3223*(2).

Bouton, M. E. (2004). Context and behavioral processes in extinction. Learning & Memory (Cold Spring Harbor, N.Y.), 11(5), 485 –494. https://doi.org/10.1101/lm.78804

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

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystkowski, J., Chowdhury, N., & Baker, A. (2008). Optimizing inhibitory learning during exposure therapy. Behaviour Research and Therapy, *46*, 5–27. http://doi.org/10.1016/j.brat.2007.10.003

Craske, M. G., & Mystkowski, J. L., (1999). Exposure therapy and extinction: Clinical studies. Fear and Learning: From Basic Processes to Clinical Implications. American Psychological Association, Washington. http://dx.doi.org/10.1037/11474-011. pp. 217–233.

Craske, M. G., & Mystkowski, J. L, (2006). Exposure therapy and extinction: clinical studies. In: Craske, MG.; Hermans, D.; Vansteenwegen, D., editors. Fear and Learning: Basic Science to Clinical Application. Washington, DC: APA Books; 2006.

Craske, M. G., & Stein, M. B. (2016). Anxiety. *Lancet* http://doi.org/10.1016/S0140-6736(16)30381-6

Craske, M. G., Stein, M. B., Eley, T. C., & Milad, M. R. (2017). Anxiety disorders. *Nature Publishing Group*, *3*(May), 1–18. http://doi.org/10.1038/nrdp.2017.24

Craske, M. G., Treanor, M., Conway, C., Zbozinek, T., & Vervliet, B. (2014). Maximizing Exposure Therapy : An Inhibitory Learning Approach, (310), 10–23. http://doi.org/10.1016/j.brat.2014.04.006.

Curtis, G. C., Magee, W. J., Eaton, W. W., Wittchen, H.-U., & Kessler, R. C. (1998). Specific fears and phobias. Epidemiology and classification. The British Journal of Psychiatry, 173(3), 212. http://doi.org/10.1192/bjp.173.3.212

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. H. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry*, *15*(3), 245–258. http://doi.org/10.1002/wps.20346

Davis III, T. E., Ollendick, T. H., & Öst, L.-G. (2009). Intensive Treatment of Specific Phobias in Children and Adolescents. *Cognitive and Behavioral Practice*, *16*(3), 294–303. http://doi.org/10.1016/j.cbpra.2008.12.008

Davey, G. C. L. (1992). Classical conditioning and the acquisition of human fears and phobias: A review and synthesis of the literature. Advances in Behaviour Research and Therapy, 14(1), 29–66. http://doi.org/10.1016/0146-6402(92)90010-L

Deak, S. & Kristoffersson, G. (2016) 16). Rädslan för det som finns och inte finns – En randomiserad kontrollerad jämförelse av utfall mellan sedvanlig ensessionsbehandling och behandling med virtuella stimuli mot spindelfobi. Psykologexamensuppsats, 30 hp, Stockholms Universitet, Stockholm.

Dafgård, P. (2017). Overklig rädsla. En jämförelse av långtidseffekterna hos Östs ensessions-behandling och virtual reality-behandling vid spindelfobi. Psykologexamensuppsats, 30 hp, Stockholms Universitet, Stockholm.

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., Hermans, D. (2015). Fear generalization in humans: systematic review and implications for anxiety disorder research. Behav. Ther. 46, 561–582. http://dx.doi.org/10.1016/j.beth.2014.10.001.

Eysenck, H. J. (1979). The conditioning model of neurosis. Behavioral and Brain Sciences, 2(2), 155−166.

Field, A. P. (2006). Is conditioning a useful framework for understanding the development and treatment of phobias?, *26*, 857–875. http://doi.org/10.1016/j.cpr.2005.05.010

Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics*. (5. ed.) London: Sage Publications.

Fisher, R. J., (1993). Social Desirability Bias and the Validity of Indirect Questioning, *Journal of Consumer Research*, *20*(2), 303–315, https://doi.org/10.1086/209351

Fitzgerald, P. J., Seemann, J. R., & Maren, S. (2014). Can fear extinction be enhanced ? A review of pharmacological and behavioral findings. *Brain Research Bulletin*, *105*, 46–60. http://doi.org/10.1016/j.brainresbull.2013.12.007

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*(1), 33–39. http://doi.org/https://doi.org/10.1016/0005-7967(95)00048-3

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. Psychological Bulletin, 99 (1), 20 – 3 5. http://doi.org/10.1037/0033 - 2909.99.1.20

Foa, E. B., & McNally, R. J. (1996). Mechanisms of change in exposure therapy. In M. (Ed.), Current controversies in the anxiety disorders (pp. 329–343). New York: The Guilford Press.

Ginsburg, G. S., Becker, E. M., Keeton, C. P., Sakolsky, D., Piacentini, J., Albano, A. M., Compton, S. N., Iyengar, S., Sullivan, K., Caporino, N., Peris, T., Birmaher, B., Rynn, M., March, J., Kendall, P. C., (2014). Naturalistic follow-up of youths treated for pediatric anxiety disorders. JAMA Psychiatry 71, 310. http://dx.doi.org/10.1001/ jamapsychiatry.2013.4186.

Hellström, K., Fellenius, J., & Öst, L. G. (1996). One versus five sessions of applied tension in the treatment of blood phobia. Behaviour Research and Therapy, 34(2), 101−112.

Hellström, K., & Öst, L.-G. (1995). One-session therapist directed exposure vs two forms of manual directed self-exposure in the treatment of spider phobia. *Pergamon, 33*(8), 959–965.

Hermans, D., Craske, M. G., Mineka, S., & Lovibond, P. F. (2006). Extinction in Human Fear Conditioning. http://doi.org/10.1016/j.biopsych.2005.10.006

Hermans, D., Vansteenwegen, D., Craske, M.G., (2006). Fear and Learning: From Basic Processes to Clinical Implications. American Psychological Association, Washington, DC.

Hesser, H. (2015). Modeling individual differences in randomized experiments using growth models: Recommendations for design, statistical analysis and reporting of results of internet interventions. *Internet Interventions*, *2*(2), 110–120. http://doi.org/10.1016/j.invent.2015.02.003

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.

Jacobson, N. S., Follette, W. C. & Revenstorf, D. (1984). Psychotherapy outcome research: methods for reporting variability and evaluating significance. Behavior Therapy, IS, 336-352.

Jacobson, N. S., & Truax, P. (1991). Clinical Significance : A Statistical Approach to Defining Meaningful Change in Psychotherapy Research, *59*(1), 12–19.

James, A. C., James, G., Cowdrey, F. A., Soler, A. & Choke, A., (2013). Cognitive behavioural therapy for anxiety disorders in children and adolescents. Cochrane Database Syst. Rev. 3, CD004690.

Kessler, R. C., Angermeyer, M., Anthony, J. C., De Graaf, R., Demyttenaere, K., Gasquet, I., … For the WHO World Mental Health Survey Consortium, T. B. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*, *6*(3), 168–176.

Lang, A. J., & Craske, M. G. (2000). Manipulations of exposure-based therapy to reduce return of fear : a replication, *38*, 1–12. https://doi.org/10.1016/S0005-7967(99)00031-5

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

Matthews, A. J., Mackintosh, C., Williams, S., Williams, M., & Kirkby, K.C., (2016). Habituation of self-reported anxiety and cortical hyper-vigilance during image-based exposure to spiders. Journal of Behavior Therapy and Experimental Psychiatry, 2017-03-01, Volume 54, Pages 150-157, 2016 Elsevier Ltd. http://dx.doi.org/10.1016/j.jbtep.2016.07.012

Miloff, A., Lindner, P., Hamilton, W., Reuterskiöld, L., & Andersson, G. (2016). Single-session gamified virtual reality exposure therapy for spider phobia vs . traditional exposure therapy : study protocol for a randomized controlled non-inferiority trial. *Trials*, 1–8. http://doi.org/10.1186/s13063-016-1171-1

Mowrer, O. H. (1956). Two-factor learning theory reconsidered, with special reference to secondary reinforcement and the concept of habit. Psychological Review, 63(2), 114–128. http://doi.org/10.1037/h0040613

Muris, P., Mayer, B., & Merckelbach, H. (1998). Trait anxiety as a predictor of behaviour therapy outcome in spider phobia. Behavioural and Cognitive Psychotherapy, 26(1), 87–91. http://doi.org/10.1017/S1352465898000095

Muris, P., & Merckelbach, H. (1996). A comparison of two spider fear questionnaires. *Journal of Behavior Therapy and Experimental Psychiatry, 27*(3), 241-244. doi:10.1016/s0005-7916(96)00022-5

Myers, K. M., & Davis, M. (2007). Mechanisms of fear extinction. *Molecular Psychiatry*, *12*(2), 120–150. http://doi.org/10.1038/sj.mp.4001939

Mystkowski, J. L., Craske, M. G., & Echiverri, A. M. (2002). Treatment Context and Return of Fear in Spider Phobia. *Behavior Therapy, 33,* 399–416. https://doi.org/10.1016/S0005-7894(02)80035-1

Mystkowski, J. L., Craske, M. G., Echiverri, A. M., & Labus, J. S. (2006). Mental Reinstatement of Context and Return of Fear in Spider-Fearful Participants, *37*, 49–60.

Mystkowski, J. L., Mineka, S., Vernon, L. L., & Zinbarg, R. E. (2003). Changes in Caffeine States Enhance Return of Fear in Spider Phobia, *71*(2), 243–250. http://doi.org/10.1037/0022-006X.71.2.243

Ollendick, T. H., Raishevich, N., Davis III, T. E., Sirbu, C., & Öst, L.-G. (2010). Specific Phobia in Youth : Phenomenology and Psychological Characteristics. *Behavior Therapy*, *41*(1), 133–141. http://doi.org/10.1016/j.beth.2009.02.002

Ollendick, T. H., Öst, L.-G., Reuterskiöld, L., Costa, N., Cederlund, R., Sirbu, C., Davis III, T. E., & Jarrett, M. A. (2009). One-Session Treatment of Specific Phobias in Youth : A Randomized Clinical Trial in the United States and Sweden, *77*(3), 504–516. http://doi.org/10.1037/a0015158

Pallant, J. (2013). *SPSS Survival Manual* (5th ed.). Berkshire, England: McGraw-Hill.

Pittig, A., Treanor, M., Lebeau, R. T., & Craske, M. G. (2018). Neuroscience and Biobehavioral Reviews The role of associative fear and avoidance learning in anxiety disorders : Gaps and directions for future research, *88*(March), 117–140. http://doi.org/10.1016/j.neubiorev.2018.03.015

Rachman, S. (1977). The Conditioning Theory of Fear-Aquisition: A Critical Examination. *Behaviour Research and Therapy*, *15*, 375–387.

Rachman, S. J. (1980). Emotional processing. Behaviour Research and Therapy, 18, 51–60.

Rachman, S. J. (1989). The return of fear. Review and prospect. *Clinical Psychology Review*, *9*, 147-168.

Rapee, R. M., Schniering, C. A., Hudson, J. L., (2009). Anxiety disorders during childhood and adolescence: origins and treatment. Annu. Rev. Clin. Psychol. 5, 311–341. http://dx.doi.org/10.1146/annurev.clinpsy.032408.153628.

Rodriguez, B. I., Craske, M. G., Mineka, S., & Hladek, D., (1999). Context-specificity of relapse: effects of therapist and environmental context on return of fear. Behaviour Research and Therapy. 37:845–862.10.1016/S0005-7967(98)00106- [PubMed: 10458048]

Rowe, M. K., & Craske, M. G. (1998) Effects of varied-stimulus exposure training on fear reduction and return of fear. Behaviour research and therapy. 36,719–734. https://doi.org/10.1016/S0005-7967(97)10017-1

Salkovskis, P. M., Hackmann, A., Wells, A., Gelder, M. G., & Clark, D. M. (2007). Belief disconfirmation versus habituation approaches to situational exposure in panic disorder with agoraphobia: A pilot study. *Behaviour Research and Therapy*, *45*(5), 877–885. http://doi.org/10.1016/j.brat.2006.02.008

Socialstyrelsen. (2010). Nationella riktlinjer för vård vid depression och ångestsyndrom 2010: stöd för styrning och ledning. Stockholm: Socialstyrelsen.

Sripada, R. K., & Rauch, S. A. M. (2015). Journal of Anxiety Disorders Between-session and within-session habituation in Prolonged Exposure Therapy for posttraumatic stress disorder : A hierarchical linear modeling approach. *Journal of Anxiety Disorders*, *30*, 81–87. http://doi.org/10.1016/j.janxdis.2015.01.002

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(7), 1047–1059. http://doi.org/10.1017/S0033291707000086

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

Tasca, G. A., & Gallop, R. (2009). Multilevel modeling of longitudinal data for psychotherapy researchers: I. The basics. Psychotherapy Research, 19 (March), 429–437. https://doi.org/10.1080/10503300802641444

Thyer, B. A., Papsdorf, J. D., Davis, R., & Vallecorsa, S. (1984). Autonomic correlates of the subjective anxiety scale. Journal of Behaviour Therapy and Experimental Psychiatry, 15, 3–7.

Thyer, B. A., & Himle, J. (1987). Phobic anxiety and panic anxiety: How do they differ? Journal of Anxiety Disorders, 1, 59–67.

Tolin, D. F. (2010). Is cognitive-behavioral therapy more effective than other therapies? A meta-analytic review. Clin. Psychol. Rev. 30, 710–720. http://dx.doi.org/10.1016/j. cpr.2010.05.003.

Tsao, J. C. I., & Craske, M. G. (2000) Timing of treatment and return of fear: Effects of massed, uniform, and expanding spaced exposure schedules. Behavior therapy. 31:479–497.10.1016/ S0005-7894(00)80026-X

Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a return to the original acquistion context, *43*, 323–336. http://doi.org/10.1016/j.brat.2004.01.001

Vlaescu, G., Alasjö, A., Miloff, A., Carlbring, P., & Andersson, G. (2016). Features and functionality of the Iterapi platform for internet-based psychological treatment. Internet Interventions, 6,107-114.

Wang, P. S., Angermeyer, M., Borges, G., Bruffaerts, R., Tat Chiu, W., De Girolamo, G., … For the WHO World Mental Health Survey Consortium T. B. (2007). Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*, *6*(3), 177–185.

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. Journal of Experimental Psychology, 3,1−14.

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

Wolpe, J. (1968). *Conditional reflex : a Pavlovian Journal of Research & Therapy,* Volume 3, Issue 4, pp 234–240 3: 234. https://doi.org/10.1007/BF03000093

Zlomke, K., & Davis III, T. E. (2008). One-Session Treatment of Specific Phobias : A Detailed Description and Review of Treatment Efficacy, *39*, 207–223.

Öst, L.-G. (1987). Age of onset in different phobias. Journal of Abnormal Psychology, 96(3), 223–229. http://doi.org/10.1037/0021-843X.96.3.223

Öst, L.-G. (1989). One-session treatment for specific phobias. Behaviour Research and Therapy, 27(1), 1–7. http://doi.org/10.1016/0005-7967(89)90113-7

Öst, L.-G. (1996). One-session group treatment of spider phobia. Behaviour Research and Therapy, 34(9), 707–715. http://doi.org/10.1016/0005-7967(96)00022-8

Öst, L-G. (2010). Manual for one session treatment of specific phobias. Stockholm University

Öst, L.-G. (2013). KBT inom psykiatrin. (L. G. Öst, Ed.) (2., rev. o). Stockholm: Natur & kultur.

Öst, L.-G., Brandberg, M., & Alm, T. (1997). One versus five sessions of exposure in the treatment of flying phobia. Behaviour Research and Therapy, 35(11), 987–996. http://doi.org/10.1016/S0005-7967(97)00077-6

Öst, L.-G., Ferebee, I., & Furmark, T. (1997). One-session group therapy of spider phobia: Direct versus indirect treatments. Behaviour Research and Therapy, 35(8), 721–732.

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, 22(3), 407–422. http://doi.org/10.1016/S0005-7894(05)80374-0

Öst, L.-G., Svensson, L., Hellström, K., & Lindwall, R. (2001). One-Session Treatment of Specific Phobias in Youths : A Randomized Clinical Trial, *69*(5), 814–824. http://doi.org/10.1037//0022-006X.69.5.814