Interventions targeting negative mental imagery in social anxiety: a systematic review of characteristics and outcomes.

Interventies gericht op negatieve mentale beelden bij sociale angst: een systematische review van de kenmerken en uitkomsten.

Alcinda Costa

s3777855

Master Thesis – Clinical Psychology

13/01/2022

Department of Psychology
Faculty of Behavioural Sciences
University of Groningen

Examiner/Daily supervisor:
M. Thunnissen, M.H. Nauta, P. de Jong

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned.

Abstract

In social anxiety individuals often experience negative imagery of themselves in feared social situations. Mental images are an important factor in the maintenance of social anxiety. This systematic review aims to provide an overview of the characteristics of imagery-based interventions targeting negative mental imagery within social anxiety as well as the effect of said interventions on social anxiety symptoms and image vividness and emotionality. The research was preregistered in PROSPERO, followed by a systematic search, conducted using MEDLINE and PsychINFO, from which data was imported into Covidence, a systematic review program. Risk of bias for randomized controlled trials and quasi-experiments was assessed using appraisal checklists from the Joanna Briggs Institute (2020). N = 21 studies with a total of N = 714 participants were included in the review. The included studies showed a variation of imagery interventions: imagery rescripting (n = 12), desensitization (n = 1), eye movement desensitization and reprocessing (n = 5) and imaginal exposure (n = 3). Within-group effect sizes for vividness were all <0.2, effect sizes for emotionality were small (n = 2), medium (n = 4) and large (n = 13). Within-group effect sizes for social anxiety were small (n = 8), medium (n = 3) and large (n = 21). Between-group effects showed a variety of difference scores, mostly medium (n = 11) and large (n = 17) on either social anxiety or imagery characteristics. Taken together, imagery interventions have a reducing effect on social anxiety and on image emotionality. Due to not labelling values to the quality assessment, the included studies varied much in quality. Future research on the integration of imagery interventions within social anxiety is needed to provide clinicians with tools to reduce image emotionality and social anxiety itself.

Samenvatting

Bij sociale angst ervaren individuen vaak negatieve mentale beelden in gevreesde sociale situaties. Mentale beelden zijn een belangrijke factor bij het in stand houden van sociale angst. Deze systematische review geeft een overzicht van de kenmerken van interventies die gericht zijn op negatieve mentale beelden binnen sociale angst, evenals het effect van deze interventies op sociale angstsymptomen en levendigheid en emotionaliteit van de beelden. Het onderzoek is vooraf geregistreerd in PROSPERO, hierna is een systematische zoekopdracht in MEDLINE en PsychINFO uitgevoerd, waaruit data werden geïmporteerd in Covidence, een systematisch reviewprogramma. Risk of bias voor randomized controlled trials en quasi-experimenten werd beoordeeld via beoordelingschecklists van het Joanna Briggs Institute (2020). *N* = 21 studies met in totaal *N* = 714 deelnemers werden geïncludeerd in de review. De studies gebruikten verschillende interventies: imagery rescripting (*n* = 12), desensitiation (*n* = 1), eye moment desensitization and reprocessing (*n* = 5) en imaginaire exposure (*n* = 3). Within-group effecten voor de mentale beelden zijn berekend voor *n* = 8 studies. Within-group effectgroottes voor levendigheid van het beeld waren allemaal kleiner dan 0,2. Effectgroottes voor emotionaliteit van het beeld waren klein (*n* = 2), gemiddeld (*n* = 4) en groot (*n* = 13). Within-group effectgroottes voor sociale angst deze klein (*n* = 8), gemiddeld (*n* = 3) en groot (*n* = 21). Effecten tussen groepen lieten een verscheidenheid aan verschilscores zien: medium (*n* = 11) en groot (*n* = 17) op sociale angst- of beeldkenmerken. Beeldinterventies hebben een reducerend effect op sociale angst en op emotionele lading van het beeld. Omdat er geen waardes gebruikt zijn bij het interpreteren van de kwaliteitsbeoordeling, variëren de artikelen veel in hun kwaliteit. Er is meer onderzoek nodig om beeldinterventies binnen sociale angst behandelingen te integreren om sociale angst te verminderen evenals de emotionele lading van de beelden.

Table of contents

Individuals with social anxiety experience intense fear of situations where others might evaluate them in a negative way (Fehm et al., 2005). Social anxiety has a 12-month prevalence of 2.0% for matching the criteria of social anxiety disorder according to the DSM-IV, whereas the prevalence of subclinical social anxiety is between 3.0 and 7.5% (Fehm et al., 2005). Social anxiety has a large impact on individuals, preventing them to participate in social events and causing problems in their school- or professional setting, including school refusal and an increased risk of dropping out of school (Stein & Stein, 2008). Social anxiety can lead to avoidance of situations where these social judgements might occur (Stein & Stein, 2008). Avoidance behaviours, such as not participating in events that are feared, not going to public places, not presenting something in front of a large group, are done because individuals with social anxiety fear that they will be evaluated negatively by others. People can also show safety behaviours, which are behaviours where individuals go near a feared situation but behave in such a way that they do not participate fully, for instance by speaking softly, staying in the background and speaking as less as possible and only speaking when asked a question as opposed to starting a conversation (Pittig et al., 2020). However, to overcome fear, one needs positive experiences of the feared event, including the experience that the anticipated catastrophe did not happen, and the experience that one can cope with negative events (Wilkins & Wallace, 1998). By avoiding all experiences, individuals also avoid possible positive experiences.

In the cognitive model of anxiety, Clark and Wells (1995) describe that negative mental imagery can play a big role in the maintenance of social anxiety. These negative mental images are often about feared social situations. These images are processed in much detail and are therefore strongly encoded in memory. Individuals with social anxiety believe the negative mental self-images to be true, making it more likely for them to avoid the situations because they expect negative outcomes to be likely (Clark & Wells, 1995). For instance, when someone has a fear of blushing in

social situations, the image they see in their head can show them excessively blushing, from which they tend to believe that that will happen in the actual situation and therefore avoid it. These negative mental images may increase and maintain the fear within social anxiety, which may lead to avoidance behaviours with regards to the feared event, thus maintaining the social anxiety disorder? (Clark & Wells, 1995). People that report seeing these images clearly experience more unpleasant bodily activity than people that reported a poor mental image, indicating that the vividness of the mental imagery might be correlated with the amount of felt bodily symptoms of fear (Holmes & Matthews, 2005). This is why targeting negative mental imagery in individuals with social anxiety is important to look into.

Because negative images in social anxiety seem to have a large influence on avoidance behaviours and on the level of fear, it is important look into how negative mental imagery can be targeted in social anxiety. Several cognitive behavioural treatments for social phobia also have techniques to correct distorted self-images, they state it is important to help patients become aware of these processes and the way in which they maintain the problem (Clark & Wells, 1995). However, less is known about treatments that explicitly target the negative mental imagery that play a role in social anxiety. Some interventions have been developed to specifically target negative imagery, with the intention to achieve symptom reduction. Research shows that reducing the vividness of the mental image and thereby reducing the emotional response can decrease avoidance and safety behaviours (Bisson et al., 2007). Eye movement desensitisation and reprocessing (EMDR), imagery rescripting (ImRe) and imaginal exposure (IE) are interventions that have been proven to be successful in reducing the vividness of a negative mental image, which can reduce the level of emotions or distress gained from the mental images. These imagery-based interventions have successfully been applied with traumatic images within post-traumatic stress disorders (Bisson et al., 2007; Arntz et al., 2012).

The current systematic review and meta-analysis is aimed at providing more information about interventions targeting negative mental imagery in social anxiety, and is thereby looking into the characteristics of found interventions such as the number of sessions, used therapeutic technique to target the images, the effect of the intervention on social anxiety symptoms and the vividness and emotionality of the image itself, together answering the following research question: 'What are the characteristics and outcomes of imagery-based interventions in social anxiety?'

Method

Search procedure

The plan for this research was to be preregistered in Prospero (see Appendix 2). In this review, the PRISMA guidelines were used as well (Page et al., 2020). We searched PsychINFO and MEDLINE databases on the 23rd of July 2021. Language was restricted to English. There were no restrictions for publication year. Based on our research question, the following combination of search terms was used: "social anxiety disorder" or "social phobi*" or "social* anxi*" or "SAD" or "performance anxiety" or "speech anxiety" or "speaking anxiety" AND "imagery" or "mental imag*" or "negative imag*" or "negative memor*" or "imagery-based" AND " intervention" or "training" or "therapy" or "imag* exposure" or "imagery rescripting" or "ImRs" or "imag* morphing" or "emotive imagery" or "EMDR" or "eye movement desensitization therapy" or "eye movement desensitization and reprocessing" or "dual-task*" or "competing task". A program, Covidence, was then used to conduct the review process. Covidence is an online screening and data extraction tool for researchers to conduct (systematic) reviews and meta-analysis (Covidence systematic review software, www.covidence.org). Here, the articles were imported and screened.

Inclusion and exclusion criteria

Studies were included when 1) regarding an intervention targeting negative mental images in social anxiety, 2) defined as an imagery-based intervention that focuses on reducing (the impact

of) relevant negative mental images within social anxiety, 3) the sample had either clinical or subclinical social anxiety, 4) the study reported outcome measurement of social anxiety and/or imagery characteristics and the effect of the intervention or 5) if the studies were randomized control trials, quasi-experiments or case studies.

Studies were excluded when 1) the study did not concern an intervention that focused mainly (less than 50% of the intervention) on negative mental imagery in social anxiety, 2) the intervention focused on positive imagery in social anxiety, 3) the participants had no social anxiety, 4) the study focused on inducing rather than reducing (the impact of) negative mental images, 5) there were no outcome measurements on both the image characteristics or social anxiety, 6) they were (systematic) reviews or dissertations, 7) the study concerned individuals with another primary diagnosis than social anxiety or 8) the study concerned subclinical symptoms of a condition other than social anxiety. No restrictions were made with regards to age or gender.

Study selection

The study selection process was done via the following steps. The search was done using the search terms as listed above. The articles that the search provided were imported in a data extraction program. Then, two reviewers independently reviewed the articles starting with reading the abstracts as a first selection stage using the inclusion and exclusion criteria as listed above as far as abstracts provided such information. Then, accepted articles were screened based on their full text using the inclusion and exclusion criteria. Any disagreements in the selecting of studies between the reviewers were resolved using the conflict section in Covidence. Articles that ended up in the conflict section were re-read by all reviewers and via a meeting between reviewers, a mutual decision about the selection of such articles was made.

Data extraction

The following data was extracted from the selected studies: study design (including use of control group), sample size (if applicable, sample size for the control condition), average age of sample, clinical or subclinical state of sample, whether the ages of the sample are below 18, characteristics of how social anxiety was measured in the sample, characteristics of how the imagery was measured in the sample, characteristics of the intervention (type, form, imagery targets, amount of sessions, duration of session), effects of the intervention (mean, standard deviation, effect size and pre- and post-measurement of social anxiety, mean, standard deviation, effect size of imagery vividness and emotionality at pre- and post-intervention measurement and amount of dropout cases from interventions). Note that in this review, an intervention was seen as control group when the intervention did not target negative social anxiety images, and if the article also has an intervention that did target negative social anxiety images.

Primary outcomes

The primary outcomes of the data extraction were the pre- and post-measurement of social anxiety and the measured effect of the imagery-intervention on the vividness and emotionality of the image as well as on the measurement of social anxiety. The data extraction was done through a data extraction form, from which the outcome tables were presented in the chapter ‘Results’. Unpublished studies weren’t included in the search.

Risk of bias assessment

The risk of bias for each study was assessed independently by two reviewers of the team using two appraisal tools from the Joanna Briggs Institute, the Checklist for Quasi-Experimental Studies and Checklist for Randomized Controlled Trials (Joanna Briggs Institute, 2017). The complete checklists can be found in Appendix 2.

Statistical analysis

The within-group effect as well as the between-group effect was calculated, using a random effects model. For studies using only an intervention condition and no control condition, the within-group effect was calculated for pre- and post-intervention characteristics for both social anxiety- and imagery characteristics. For studies using an intervention condition as well as a control condition, both the within-group effect and the between-group effect were calculated for social anxiety- and imagery characteristics. For the within-group effect, Cohen's d formula was used with $d = \frac{\text{Mpost}-\text{Mpre}}{\text{s}}$ where Mpost is the mean of the measurements of social anxiety characteristics or imagery characteristics post-intervention and Mpre is the mean of the measurements of social anxiety characteristics or imagery characteristics pre-intervention, and s is the standard deviation of the pre-intervention measurements of the intervention group. For the between-group effect, the formula $d = \text{d}control - dintervention$ was used to state the difference between the effect sizes of the control group and the intervention group, where d*control* is the effect size of the control condition and d*intervention* is the effect size of the intervention condition, and s = the standard deviation of the pre-intervention measurements of the intervention group. The between-group effects were be displayed in a separate table from the within-group effects.

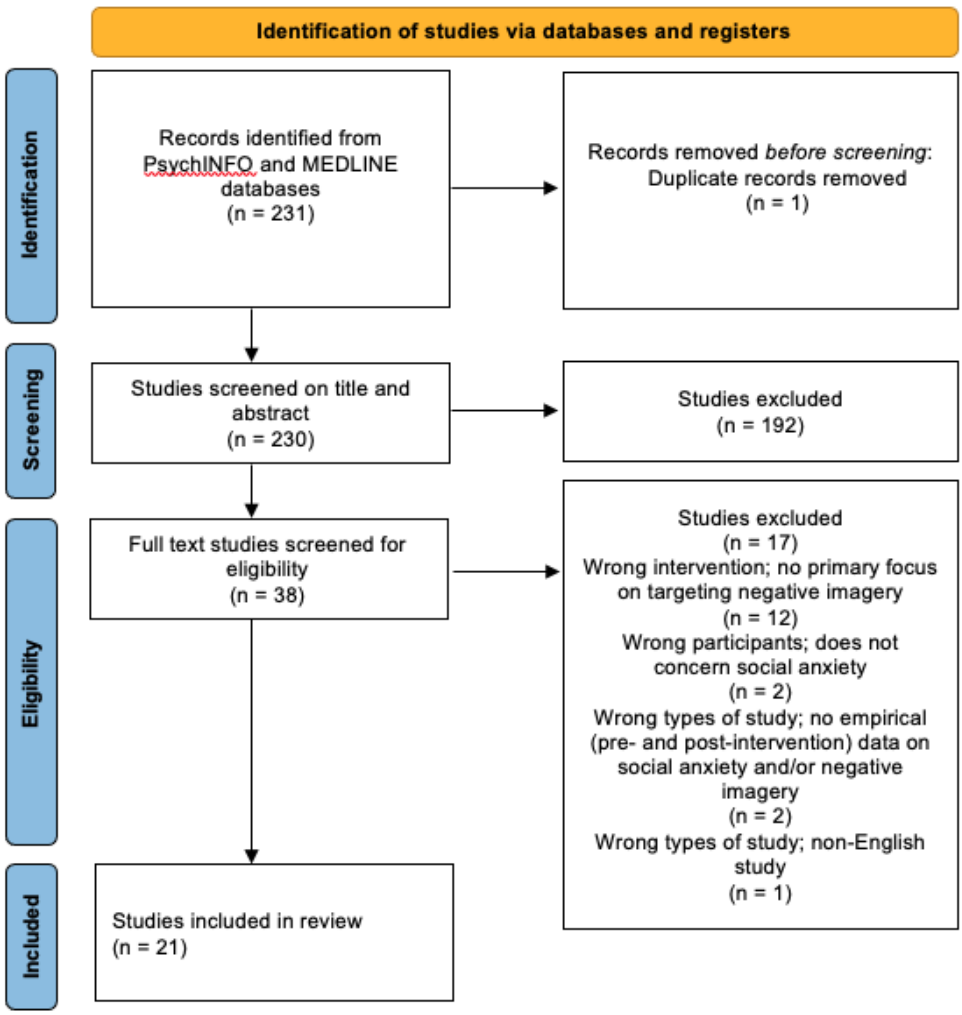
Results

Study characteristics

A PRISMA diagram summarizing the search and selection process is displayed in Figure 1 below. We identified 230 individual studies, of which 21 met the inclusion criteria and were included in the data extraction, which are described in Table 1. From all included studies (N = 21), the majority were randomized controlled trials (n = 15). The average age was above 18 in all included studies. Of all included studies (N = 21), there were studies with participants with clinical anxiety (n = 13; N = 326) and subclinical (n = 8; N = 403) social anxiety. The number of participants in the

clinical samples range from *n* = 6 until *n* = 60. The number of participants in the subclinical samples range from *n* = 27 until *n* = 71. The total number of participants from all 21 selected studies of intervention groups as well as control groups (when applicable) are N = 714, with a range from *n* = 6 until *n* = 71 (median; Q2 = 29).

Figure 1



PRISMA Flow Diagram of study selection (Page et al., 2021).

Intervention characteristics

There was a variation of imagery interventions used in the included studies: imagery rescripting (n = 12), desensitization (n = 1), eye movement desensitization and reprocessing (n = 5) and imaginal exposure (n = 3) (see Table 1). Most imagery interventions were therapist-guided (n = 16) whilst some were computerized tasks (n = 5). From all EMDR interventions, one was script driven. Most interventions consisted of one session (n = 14) and some consisted of multiple sessions (n = 7, range 2-17 sessions). Durations of interventions differed. For imagery rescripting interventions, the duration ranged from 30-45 min (*n = 2)* to 60-90 min *(n = 1)* to 90 min or longer *(n = 7*. The duration of EMDR-interventions ranged from 60 seconds *(n = 2)* to 2-3 minutes *(n = 2)* to 90 minutes *(n = 1)*.

The types of images targeted in the intervention were autobiographical memories (*n* = 14), script driven images (*n* = 3), a problematic recent social anxiety event (*n* = 1), a flashforward image regarding social anxiety (*n* = 1) a public speaking situation (*n* = 1) and an intrusive social anxiety image (*n* = 1). From all twenty-one included studies, fourteen studies included either a control group (*n* =10) and/or a control period (*n* = 4). Eight studies had no control group (*n* = 8). Note that some studies had multiple control conditions per study. Control conditions existed of no-treatment control (*n* = 2), verbal restructuring (*n* = 1), eye movements with neutral imagery without desensitization and reprocessing (*n* = 3), group cognitive behavioural therapy (*n* = 1), cognitive modification (*n* = 1), in vivo exposure (*n* = 1), waitlist control (*n* = 1), exercise condition (*n* = 1), supportive counselling (*n* = 1), a control period of 3 weeks (*n* = 1) and vicarious video treatments (desensitization, *n* = 1, desensitization with coping imagery, *n* = 1, and cognitive modification, *n* = 1). From all included studies, five reported follow-up measurements.

Intervention effects

Social anxiety characteristics

The outcomes in this systematic review were pre- and post-interventions measurements of the mean and standard deviation of imagery measurements (see Table 2) and social anxiety measurements (see Table 3). Of all 21 included studies, 12 studies reported pre- and post-measurements of social anxiety. Social anxiety was mostly measured with the Personal Report of Confidence as a Speaker questionnaire, a twelve-item self-report measure by Paul (1996) used by Weissberg et al. (1977), Schwartz & Kaloupek (1987) and Homer et al. (2016), (Brief) Fear of Negative Evaluation, a self-report questionnaire (Frets, Kevenaar & van der Heiden, 2014; Knutsson et al., 2020; Nilsson, Eriksson & Jarild, 2019; Lee, 2013; Norton & Abbott, 2016; Norton Abbott, Dobinson, Pepper & Guastella, 2021; Takanashi, Yoshinaga, Oshiro, Matsuki, Tanaka, Ibuki, Oshima, Urao, Matsuzawa & Shimizu, 2020; Wild, Hackmann & Clark, 2007; Wild, Hackmann & Clark, 2008), the Social Interaction Anxiety Scale (Mattick & Clarke, 1998) used by Nilsson et al. (2012), Norton et al. (2016) and Norton et al. (2021). Other studies used the Subjective Units of Distress as a measurement (Carrigan & Levis, 1999; Norton et al., 2016; Rubin, Spates, Johnson & Jouppi, 2009; Vrielynck & Philippot, 2009) or physical measures such as Facial electromyography (Rubin et al., 2009) mean skin conductance level (Hyett et al., 2018) or number of skin conductance responses (Hyett et al., 2018).

Image characteristics

Pre- and post-intervention image vividness was reported by 10 of 21 included studies and pre- and post-intervention emotionality was reported by 11 of 21 studies. Imagery vividness and emotionality were measured on a Visual Analogue Scale. Some studies reported other image measurements such as core beliefs about self (true/valid) (Norton et al., 2021), feeling ashamed or proud (Reimer, 2015), image frequency (Nilsson, 2012; Wild et al., 2008) which were beyond the

scope of this review. Of all (N = 21) studies, five studies reported follow-up measurements. Although all interventions targeted a negative image related to social anxiety, many targeted a subcategory of social anxiety which the images originated from, such as public speaking anxiety or performance anxiety. Aside from two studies, all studies inclined that the targeted negative images were related to a social anxiety experience of the participant. Only two studies clearly stated that the targeted image was script driven, meaning that a script, read by the therapist provided a descripting of a social anxiety image that the participants were then told to imagine (Weissberg et al., 1977; Kearns & Engelhard, 2015).

Table 1

Study characteristics and intervention characteristics

Author (year)	Study Design	*N*	Study population	Mean (sd) age of intervention group	Intervention (*n*)	Number of sessions, duration of sessions, duration of total intervention	Targeted image	Control	Follow-up
Carrigan (1999)	RCT	71	Subclinical	Unclear	EMDR with fear-relevant image and eye movements *(18)* EMDR with fear-relevant image without eye movements *(18)*	1 session with 9 eye movement/imagery phases (90min total)	Autobiographical negative memory associated with public speaking anxiety	Control group	-
Engelhard (2012)	Within-subjects experimental design	29	Subclinical	23 (4.70)	Imagery with eye movements (EMDR) and imagery without eye movements *(29)*	1 session of six 24sec phases with imagery (2min and 24 seconds total),	Flashforward image related to performance anxiety	No control	-

Frets (2014)	Case series	6	Clinical	Age range: 21-28-30-31-40-47	Imagery rescripting *(6)*	Number of sessions ranged from 5-7 45min sessions, mean 11.2 sessions (total time ranged from 225min to 315 min)	Problematic recent social event	Control period	-
Homer (2016)	RCT	40	Subclinical	22	EMDR eye movement condition *(17)* EMDR auditory condition *(19)*	EMDR eye movement condition: 3 sessions (60sec total) EMDR auditory condition: 3 sessions (61.2 seconds total)	Autobiographical memory representing public speaking anxiety	No control	-
Homer (2018)	RCT	27	Subclinical	20 (2.51)	EMDR eye movement condition *(14)* EMDR no eye movement condition *(12)*	EM condition: 1 45min session containing 3 60sec blocks of EM No EM condition: 1 45min session containing 3 blocks of visualization (45 min total each)	Intrusive SAD image	No control	-
Hyett (2018)	RCT	58	Clinical	35.22 (14.98)	Imagery rescripting *(17)*	1 90 min session (90min total)	Autobiographical memory (negative socially evaluative situation that shaped current feelings of social situations)	Control group and waitlist control	Yes

Kearns (2015)	RCT	34	Subclinical	21.4 (2.99)	EMDR eye movement condition *(17)* EMDR no eye movement condition *(17)*	EM condition: 1 session containing 6 24sec phases of eye movement No EM condition: 1 session containing 6 24sec phases of imagery (144sec of each condition total)	Script driven flash forward image of public speaking scenario	No control	-
Knutsson (2019)	RCT	27	Clinical	25 (3.8)	Imagery rescripting *(14)*	1 90min session (90min total)	Autobiographical memory linked to negative self-imagery	Control group	4-week follow up
Lee (2013)	RCT	23	Clinical	23.92 (3.35)	Imagery rescripting *(13)*	3 sessions of 1 or 2hr (3 to 6hrs total)	Autobiographical fear-provoking memory linked to negative imagery	Control group	3-month follow-up
Nilsson (2012)	RCT	14	Clinical	36.4 (10.3)	Imagery rescripting *(7)*	1 session, duration and total time unclear	Autobiographical memory linked to SAD, linked to negative spontaneous imagery	Control group	3-week follow-up
Norton (2016)	RCT	60	Clinical	20.83 (3.99)	Imagery rescripting *(20)*	1 30-45 min session (30-45min total)	Autobiographical distressing memory linked to negative self-imagery experienced in	Control group	-

							anxiety-provoking social situations		
Norton (2021)	Case series	15	Clinical	23.7 (4.89)	Imagery rescripting *(15)*	2 45-60min sessions (90-120min total)	Autobiographical distressing memory to recurrent negative self-image in socially threatening situations	Control period	-
Reimer (2015)	RCT	25	Clinical	19.57 (1.16)	Imagery rescripting (*13)*	1 90min session (90min total)	Autobiographical memory to anxiety-provoking images of social situations	Control group	-
Romano (2020)	RCT	33	Clinical	IR: 26.18 (5.27) IE: 29.91 (12.16)	Imagery rescripting *(11)* Imaginal exposure *(11)*	1 60-90min session plus daily homework assignments about memory for 6 days (60-90min total, excluding homework time)	Autobiographical negative social memory	Control group	unclear
Rubin (2009)	RCT	39	Subclinical	21.92	Imagery exposure; Positively enhanced *(10)* Negatively supplemented dosed *(10)* Prolonged *(10)* Dosed *(9)*	1 session with 10 15sec phases of scene imagery (2.5 min total)	Script driven public speaking scene from audio recording	No control	-

Schwartz (1987)	RCT	52	Subclinical	26.4 (5.9)	Imagery exposure (unclear)	2 45min sessions containing 3 15min blocks of imagery (1.5hr total)	Public speaking situation	Control group	-
Takanashi 2020)	Case series	25	Clinical	32.08 (8.76)	Imagery rescripting *(25)*	1 or 2 90min sessions, containing 30min imagery rescripting and 30 cognitive restructuring (CR) (90-180 min total)	Autobiographical memory linked to current negative imagery	No control	-
Vrielynck (2009)	RCT	49	Subclinical	Unique/episodic group: 20.12 (5.61) Generic group: 19.54 (1.93)	Unique/episodic imaginary exposure *(25)* Generic imaginary exposure *(24)*	2 sessions (1 rehearsal session) with 15min exposure per session (15-30min total)	Autobiographical (recent) distressing memory that was considered a failure in life	No control	-
Weissberg (1977)	RCT	62	Subclinical	18.5	Desensitization *(10)* Desensitization with coping imagery *(10)*	Desensitization: 3 2hr sessions (6hr total) Desensitization with coping imagery: 3 2hr sessions (6hr total)	Script-driven scenes from a 12-item group hierarchy	Control group	Yes
Wild (2007)	Case series	14	Clinical	28.64 (3.75)	Imagery rescripting *(14)*	1 75min session containing of 30min cognitive restructuring	Autobiographical memory linked to negative image	No control	-

						and 45min imagery rescripting (75min total)			
Wild (2008)	Case series	11	Clinical	35.18 (9.36)	Imagery rescripting *(11)*	1 90min session containing 45min restructuring and 45min rescripting (90min total)	Autobiographical memory linked to recurrent negative image in social situations	Control period	-

Abbreviations: randomized controlled trial (RCT), eye movement desensitization and reprocessing (EMDR), imagery rescripting (IR), imaginal exposure (IE), cognitive restructuring (CR), not applicable (-), minutes (min), seconds (sec), hour (hr), unclear (-).

Effect sizes

Within-group effect sizes for imagery measurements were calculated using Cohen's d for n = 8 conditions in four studies (see Table 3). The effect sizes for image vividness for the intervention group was calculated for *n* = 8 conditions in *n* = 4 studies (Engelhard, Sijbrandij, van den Hout, Rutherford, Rahim & Kocak 2021; Lee & Kwon, 2013; Norton, 2016; Nilsson, Lundh & Viborg, 2012) and included small (*n* = 3), medium (*n* = 2) and large (*n* = 3) effect sizes, all in negative numbers, indicating that the image vividness was reduced. The effect sizes for image emotionality of the intervention group were calculated for *n* = 14 conditions from *n* = 11 studies (Engelhard et al., 2012; Homer, Deeprose & Andrade 2016; Kearns & Engelhard, 2015; Wild et al., 2008; Norton, 2021; Lee & Kwon, 2013; Norton, 2016; Reimer & Moscovitsch, 2015; Homer & Deeprose, 2018; Nilsson, 2012; Wild, 2007). These effect size measures included small (*n* = 2), medium (*n* = 4) and large (*n* = 13). The effect sizes were more often medium to large when the intervention concerned imagery rescripting (*n* = 9). The effect sizes for image vividness of the intervention group were calculated for *n* = 4 conditions in *n* = 4 studies. All calculated effect sizes of the intervention group for vividness were smaller than 0.2.

Within-group effect sizes for social anxiety measurements were calculated using Cohen's d for n = 13 studies (Weissberg, 1977; Schwartz & Kaloupek, 1987; Hyett et al., 2018; Homer et al., 2016; Wild et al., 2008; Norton, 2021; Lee & Kwon, 2013; Norton, 2016; Reimer & Moscovitsch, 2015; Homer & Deeprose, 2018; Knutsson et al., 2019; Nilsson, 2012; Wild et al., 2007). The effect sizes for the intervention groups included small (*n* = 13), medium (*n* = 7) and large (*n* = 20) effect sizes (note that per study, multiple measures were reported for social anxiety, thus giving multiple effect sizes per article). Almost all effect sizes were negative, indicating a reduction of (sub-)clinical social anxiety. Five studies with an imagery rescripting intervention reported large effect sizes in social anxiety measurements

in the intervention group (Hyett et al., 2018; Lee & Kwon, 2013; Nilsson, 2012; 2013; Reimer, 2015; Wild et al., 2008), as opposed to only one study using EMDR showing a large effect size in social anxiety measures (Homer & Deeprose, 2018). Of the imagery exposure studies, an effect size could only be calculated for one study. These effect sizes calculated for this study were medium to large (Schwartz & Kaloupek, 1987).

For the control group the within-group effect sizes for social anxiety were small (n = 8), medium (n = 3) and large (n = 21). Note that some studies used multiple conditions, from which multiple effect sizes were calculated. Interestingly, Weissberg (1977), reported large reductions in post-measurements of social anxiety as opposed to pre-measurements of social anxiety for the intervention group as well as the control group. Weissberg (1977) used five different control groups, namely cognitive modification, vicariously desensitization, vicariously desensitization with coping imagery, vicariously cognitive modification, exercise control group and a no treatment control. Unfortunately, the no treatment control group had no pre- or post-intervention measurements, which makes it difficult to distinguish between the effect of the used treatment, desensitization, and the control groups who also had interventions related to social anxiety. The largest intervention group effect size was with imaginal exposure combined with exercise (N = 52) as intervention (Schwartz & Kaloupek, 1987) of d = -5.041 as the effect size of heart rate during speech with speech anxiety individuals. However, the exercise control group of Schwartz & Kaloupek (1987) without imaginal exposure, also showed a large effect size in speech heart rate (d = 5.851).

The between-group effect sizes were calculated to state the overall controlled effect size (see Table 4). This was calculated only for studies with a pre- and post-intervention measurements of the control condition of either social anxiety measurements or imagery characteristics. Those studies (n = 12) are shown in Table 4. With the calculation of the

between-group effect, large outcomes were found. Because the intervention effect size was subtracted from the control effect size, large negative outcome suggest that the intervention had a larger effect on the intervention group than on the control group/period. Results showed a variety of difference scores, mostly medium (n = 11) and large (n = 17) on either social anxiety or imagery characteristics. Note that per study, sometimes multiple effect sizes were calculated when there were multiple groups or conditions to compare. A number of studies including imagery rescripting resulted in medium to large difference scores (Lee, 2013; Hyett, 2018; Knutsson, 2019; Nilsson, 2012; Norton, 2021; Reimer, 2015; Wiessberg,1977 and Wild, 2008). One study with imagery exposure showed large differences between control condition and intervention as well. Weissberg (1977), using desensitization and desensitization with coping imagery as intervention, also showed medium to large differences between effect sizes of the control condition(s) versus the intervention(s). The results on differences between the effect sizes of the control condition and intervention when looking at image vividness and emotionality show medium (Norton, 2016) and large differences (Lee, 2013; Nilsson, 2012; Reimer, 2015). These are less medium and large effects than the ones on social anxiety measurements. In studies where effect size differences were calculated, emotionality almost always showed more difference than vividness (Engelhard, 2012; Lee, 2013; Nilsson, 2012; Norton, 2016). This indicates that the interventions had a larger effect on reducing the emotionality of the image than the vividness of the image.

Table 3

Imagery Measurements of included studies

Study characteristics			Image measurements intervention group			Image measurements control group or period?			
Author (year)	*Total (N)*	Intervention (*n*)	Pre-intervention	Post-intervention	Effect size (Cohen's d)	Control *(n)*	Pre-intervention	Post-intervention	Effect size (Cohen's d)
Carrigan (1999)	71	EMDR with fear-relevant image and eye movements *(18)* EMDR with fear-relevant image without eye movements *(18)*	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-	Eye movements and relaxing imagery *(18)* Relaxing imagery only *(17)*	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-
Engelhard (2012)	29	Imagery with eye movements and imagery while looking at stationary	Vividness: 76.66 (11.23) Emotionality: 69.77 (19.57)	Vividness: 78.60 (13.14) Emotionality: 59.35 (21.20)	Vividness: 0.173 Emotionality: -0.532	-	Vividness: 75.73 (12.33) Emotionality: 68.97 (13.57)	Vividness: 78.60 (13.14) Emotionality: 66.95 (19.22)	Vividness: 0.233 Emotionality: -0.149

		circle without eye movements *(29)*							
Frets (2014)	6	Imagery rescripting *(6)*	Unclear	Vividness: unclear Emotionality: unclear	unclear	Control period: 3 week no-treatment baseline	Vividness: unclear Emotionality: unclear	Vividness: unclear Emotionality: unclear	unclear
Homer (2016)	40	EMDR eye movement condition *(17)* EMDR auditory condition *(19)*	Vividness: Eye movement condition: 77.85 (10.74) Auditory condition: 79.13 (12.58) Emotionality: Eye movement condition: 65.56 (21.80) Auditory condition: 61.55 (25.13)	Vividness: - Emotionality: Eye movement condition: 53.88 (20.74) Auditory condition: 62.11 (23.27)	Vividness: - Emotionality (EM): -0.536 Emotionality (Aud.): 0.022	-	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-
Homer (2018)	27	EMDR eye movement condition *(14)*	Vividness EM condition: 7.36 (1.69)	Vividness: - Emotionality: EM condition: 4.93 (1.94)	Vividness: - Emotionality:	-	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-

		EMDR no eye movement condition *(12)*	no-EM condition: 5.33 (2.42) Emotionality EM condition: 5.57 (1.45) No-EM condition: 5.92 (2.02)	no-EM condition: 5.67 (1.50)	EM: -0.441 No EM: -0.124				
Hyett (2018)	58	Imagery rescripting *(17)*	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-	Verbal restructuring *(22)* Waitlist control *(19)*	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-
Kearns (2015)	34	EMDR eye movement condition *(17)* EMDR no eye movement condition *(17)*	Vividness: EM condition: unclear no-EM condition: unclear Emotionality EM condition: 2.03 (2.02) Emotionality: unclear Heart rate: EM-condition:	Vividness: - Emotional intensity EM condition: unclear no-EM condition: unclear HR EM condition: 0.77 (2.48) no-EM condition: 2.50 (3.86)	HR EM: -0.624 HR no EM: 0.239	-	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-

			2.03 (2.02) No-EM condition: 1.71 (3.30)						
Knutsson (2019)	27	Imagery rescripting *(14)*	Vividness: 61.25 (29.93) Emotionality: 52.50 (27.60)	Vividness: - Emotionality: - Post-intervention Imagery distress 40.42 (20.50) 14-day follow-up Imagery distress 29.17 (22.55)	-	In vivo exposure *(13)*	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-
Lee (2013)	23	Imagery rescripting *(13)*	Vividness: 84.23 (9.54) Emotionality: 85.38 (8.02)	Vividness: 65.60 (22.17) Emotionality: 36.77 (18.39)	Vividness: -1.953 Emotionality: -6.061	First session similar to intervention group, second and third session was an attention-placebo with therapeutic attention and support *(10)*	Vividness: 79.70 (14.41) Emotionality: 82.00 (7.15)	Vividness: 65.50 (22.17) Emotionality: 66.50 (18.86)	Vividness: -0.990 Emotionality: -2.168
Nilsson (2012)	14	Imagery rescripting *(7)*	Vividness: 77.14 (18)	Vividness: 67.14 (24.3)	Vividness: -0.556	Reading task from self-help book on	Vividness: 65.71 (22.25)	Vividness: 67.14 (24.3)	Vividness: 0.064

			Emotionality: 80.00 (14.14)	Emotionality: 51.43 (26.10)	Emotionality: -2.021	cognitive behavioral therapy *(7)*	Emotionality: 68.57 (24.79)	Emotionality: 55.71 (18.13)	Emotionality: -0.519
Norton (2016)	60	Imagery rescripting *(20)*	Vividness: 22.01 (29.19) Emotionality: 13.03 (17.86)	Vividness: cognitive restructuring: 15.14 (15.93) Emotionality: imagery distress: 4.60 (6.42)	Vividness: -0.235 Emotionality: -0.471	Cognitive restructuring *(10)* Control with puzzle task *(20)*	Vividness: cognitive restructuring: 16.47 (15.55) control: 17.78 (18.77) Emotionality: imagery distress: 12.76 (13.22)	Vividness: cognitive restructuring: 15.14 (15.93) control: 16.30 (18.64) Emotionality: imagery distress: 12.96 (18.04) control imagery distress: 12.10 (16.84)	CR: Vividness: -0.086 CR: Emotionality: 0.015 Control Vividness: -0.011 Emotionality: -0.050
Norton (2021)	15	Imagery rescripting *(15)*	Vividness: - Emotionality of self-image (positive): 1.33 (0.82) (negative): 3.73 (0.80) (intensity): 2.80	Vividness: - Emotionality/impact of self-image Positive 2.00 (1.19) Negative 3.00 (1.13) Intensity 3.00 (.86) Global perceptions of self-image	Vividness: - Emotionality/impact of self-image: Positive: 0.817 Negative: -0.913 Intensity: 0.233	Control period: Group cognitive behavioral therapy *(15)*	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-

			(0.86) embarrassed/ashamed: 4.07 (0.70) Pleased/proud: 1.13 (0.35)	Embarrassed/Ashamed 2.93 (1.39) Pleased/Proud 2.00 (1.20) Emotional impact of memory Positive 2.27 (1.16) Negative 2.87 (1.06) Intensity 2.93 (.96) Global perceptions of memory Embarrassed/Ashamed 2.53 (1.30) Pleased/Proud 2.40 (1.35)	Embarrassed/ashamed: -1.629 Pleased/proud: 3.629			
Reimer (2015)	25	Imagery rescripting (*13)*	Vividness: - Emotional impact (positive): 1.23 (0.44) (negative): 4.15 (0.69) (intense): 3.69 (1.03)	Vividness: - Emotional impact positive: 2.31 (1.11) negative: 2.69 (.85) intense: 2.23 (.93)	Vividness: - Emotional impact: positive: 2.455 negative: -2.116 intense: 0.587	No-intervention control *(12)*	Vividness: - Emotional impact positive: 1.09 (0.30) negative: 4.55 (0.69) intense: 3.45 (1.04)	Vividness: - Emotional impact positive: 1.09 (0.30) negative: 4.00 (0.77) intense: 3.00 (1.26)	Vividness: - Emotional impact positive: 0 negative: -0.797 intense: -0.433

Romano (2020)	33	Imagery rescripting *(11)* Imaginal exposure *(11)*	IR: 4.55 (0.52) IE: 4.18 (0.87) Emotionality: IR: 25.27 (8.27) IE: 21.00 (7.20)	Vividness: unclear Emotionality: IR: negative affect: unclear IE: negative affect: unclear		Supportive counseling *(11)*	Vividness: 4.18 (0.87) Emotionality/negative affect: 26.55 (7.53)	Vividness: unclear Emotionality: unclear	unclear
Rubin (2009)	39	Imagery exposure; Positively enhanced *(10)* Negatively supplemented dosed *(10)* Prolonged *(10)* Dosed *(9)*	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-	-	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-
Schwartz (1987)	52	Imagery exposure (unclear)	Vividness: unclear Emotionality: unclear	Vividness: unclear Emotionality: unclear	unclear	Exercise condition: neutral imagery (unclear) Placebo control: neutral imagery (unclear)	Vividness: unclear Emotionality: -	Vividness: unclear Emotionality: unclear	unclear

Takanashi 2020)	25	Imagery rescripting *(25)*	Vividness: median = 70 (25-100) memory vividness: median = 80 (40-100) Image distress median: 70 (20-100) Memory distress median: 70 (0-100)	Vividness: - Emotionality: image distress: median = 35 (0-90) memory distress: median = 40 (0-100)	-	-	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-
Vrielynck (2009)	49	Unique/episodic imaginary exposure *(25)* Generic imaginary exposure *(24)*	Vividness: Unique/episodic: unclear Generic: unclear Emotionality: unclear	Vividness: - Emotionality: -	-	-	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-
Weissberg (1977)	62	Desensitization *(10)* Desensitization with coping imagery *(10)*	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-	Cognitive modification *(10)* Vicarious desensitization *(10)*	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-

						Vicarious desensitization with imagery *(11)* Vicarious cognitive modification *(12)* No treatment control (unclear)				
Wild (2007)	14	Imagery rescripting *(14)*	image vividness: 60.91 (27.73) memory vividness: 68.64 (24.09) Emotionality: Image distress: 54.29 (24.08) Memory distress: 66.07 (22.03)	Vividness: - Emotionality: post-rescripting image distress: 20.71 (22.69) memory distress: 23.57 (25.53) follow up (n = 8) image distress: 27.50 (26.59) memory distress: 3.75 (7.44)	Vividness: - Emotionality: image distress: -1.395 memory distress: -1.929	-	Vividness: - Emotionality: -	Vividness: - Emotionality: -	-
Wild (2008)	11	Imagery rescripting *(11)*	Vividness: pre rescripting	Emotionality: post rescripting session	Memory distress:	-	Vividness: pre control period	Vividness: post control period: -	Memory distress: -

session/follow-up after control session 51.82 (29.01) Emotionality: memory distress: 53.18 (26.48) Image distress: 47.73 (27.78)	memory distress: 8.89 (15.37) image distress: - follow-up after rescripting session memory distress: 22.27 (19.92) image distress: 19.55 (25.54)	-1.673 Image distress: -1.014	60.91 (23.75) Emotionality Memory distress: 68.18 (27.14) Image distress: 50.00 (26.55)	Emotionality: Memory distress: 60.56 (28.22) Image distress: -	0.281

Abbreviations: eye movement desensitization and reprocessing (EMDR), imagery rescripting (IR), imaginal exposure (IE), heart rate (HR), eye movement (EM), previous to (pre-), posterior to (post-), not applicable (-), auditory (aud.).

Table 2

Social anxiety measurements of included studies

Study characteristics			Social anxiety measurements intervention group				Social anxiety measurements control group		
Author (year)	*N*	Intervention	Pre-intervention	Post-intervention	Effect size (Cohen's d)	Control (*n*)	Pre-intervention	Post-intervention	Effect size (Cohen's d)
Carrigan (1999)	71	EMDR with fear-relevant image and eye movements *(18)* EMDR with fear-relevant image without eye movements *(18)*	Unclear	Unclear			unclear	unclear	unclear
Engelhard (2012)	29	Imagery with eye movements and imagery while looking at stationary circle without eye movements *(29)*	-	-	-	-	unclear	-	-

Frets (2014)	6	Imagery rescripting *(6)*	unclear	unclear	Unclear	Control period: 3 week no-treatment baseline	unclear	unclear	unclear
Homer (2016)	40	EMDR eye movement condition *(17)* EMDR auditory condition *(19)*	PRCS Eye movement condition: 40.18 (8.54) Auditory condition: 36.21 (8.27) Confidence Eye movement condition: 37.62 (19.83) Auditory condition: 36.47 (25.40) Anxiety Eye movement condition: 77.47 (13.18)	PRCS (state) Eye movement condition: 40.12 (8.85) Auditory condition: 40.00 (7.53) Confidence (state) Eye movement condition: 38.94 (21.90) Auditory condition: 37.47 (24.42) Anxiety (state) Eye movement	PRCS (state) Eye movement condition: 0.001 Auditory condition: 0.458 Confidence (state) Eye movement condition: 0.067 Auditory condition: 0.039 Anxiety (state) Eye movement condition: -0.283 Auditory condition: 0.260	-	-	-	-

			Auditory condition: 78.53 (15.41)	condition: 73.74 (14.53) Auditory condition: 82.53 (11.25)					
Homer (2018)	27	EMDR eye movement condition *(14)* EMDR no eye movement condition *(12)*	40.85 (9.43)	16.71 (0.83)	-2.560	-	-	-	-2.485
Hyett (2018)	58	Imagery rescripting *(17)*	SIPS: 30.76 (2.33)	SIPS: 27.57 (2.98)	SIPS: -1.369	Verbal restructuring *(22)* Waitlist control *(19)*	VR: SIPS: 34.64 (1.70) WC: SIPS: 4.74 (1.74)	SIPS: 33.58 (2.47)	SIPS: -0.624
Kearns (2015)	34	EMDR eye movement condition *(17)* EMDR no eye movement condition *(17)*	-	-	-	-	-	-	-
Knutsson (2019)	27	Imagery rescripting *(14)*	BFNE-II: 20.08 (6.07) LSAS-SR: 71.64 (22.07)	14-day follow-up BFNE-II 17.42 (6.75)	BFNE-II: -0.438 LSAS-SR: -0.391	In vivo exposure *(13)*	BFNE-II: 22.82 (3.49) LSAS-SR: 72.00 (29.57) Anxiety: 83.89 (15.96)	(14-day Follow-up) BFNE-II: 19.09 (6.32) LSAS-SR: 58.00 (18.04	BFNE-II: -1.069 LSAS-SR: -0.473 Anxiety: -3.028

				LSAS-SR 63.00 (27.03)				Anxiety: 35.56 (30.36)	
Lee (2013)	23	Imagery rescripting *(13)*	K-SADS: 106.54 (11.70) K-BFNE: 49.85 (3.53) LSAS-Fear: 39.38 (12.95) LSAS-Avoidance: 38.23 (13.40)	K-SADS 91.77 (7.03) K-BFNE 43.54 (4.29) LSAS-Fear 31.92 (13.88) LSAS-Avoidance 22.62 (13.84)	K-SADS: -1.262 K-BFNE: -1.788 LSAS-Fear: -0.576 LSAS-Avoidance: -1.165	First session similar to intervention group, second and third session was an attention-placebo with therapeutic attention and support *(10)*	K-SADS: 105.00 (10.04) K-BFNE: 50.10 (4.12) LSAS-Fear: 38.50 (11.15) LSAS-Avoidance: 36.00 (12.41)	K-SADS: 9.00 (7.09) K-BFNE: 48.90 (2.89) LSAS-Fear: 7.10 (9.34) LSAS-Avoidance: 34.00 (9.94)	K-SADS: -8.251 K-BFNE: -0.291 LSAS-Fear: -2.816 LSAS-Avoidance: -0.161
Nilsson (2012)	14	Imagery rescripting *(7)*	FNE: 22.71 (6.02) SIAS: 36.71 (21.96)	Follow-up 1 week later FNE: 13.43 (10.05) SIAS: 27.14 (19.23)	FNE: -1.542 SIAS: -0.436	Reading task from self-help book on cognitive	FNE: 22.14 (5.01) SIAS: 36.86 (12.03)	Follow-up 1 week later FNE: 22.71 (4.79) SIAS: 37.14 (11.51)	FNE: 0.114 SIAS: 0.023

						behavioral therapy *(7)*			
Norton (2016)	60	Imagery rescripting *(20)*	SIAS: 44.55 (12.08) B-FNE: 35.60 (7.22)	SIAS: 35.85 (12.27) B-FNE: 32.55 (7.10)	SIAS: -0.720 BFNE: -0.422	Cognitive restructuring *(10)* Control with puzzle task *(20)*	Cognitive restructuring control SIAS: 43.40 (7.97) B-FNE: 35.65 (4.90) Control SIAS: 42.70 (8.32) B-FNE: 34.90 (5.41)	Cognitive restructuring control SIAS: 34.65 (11.91) B-FNE: 29.20 (5.76) Control SIAS: 40.25 (11.17) B-FNE: 31.25 (6.53)	Cognitive restructuring control SIAS: -1.098 B-FNE: -1.316 Control SIAS: -0.294 B-FNE: -0.675
Norton (2021)	15	Imagery rescripting *(15)*	Post-GCBT: SIAS + SPS: 74.07 (25.21) BFNE: 45.71 (9.19)	Post-IR: SIAS + SPS: 73.29 (25.20) BFNE: 42.5 (9.02)	SIAS + SPS: -0.031 BFNE: -0.350	Control period: Group cognitive behavioral therapy *(15)*	Control period (pre-GCBT): SIAS + SPS: 96.23 (30.06) BFNE: 48.71 (8.51)	Control period (post-GCBT): SIAS + SPS: 74.07 (25.21) BFNE: 45.71 (9.19)	SIAS + SPS: -0.737 BFNE: -0.353
Reimer (2015)	25	Imagery rescripting (*13)*	SPIN: 41.77 (9.66) LSAS-SR: 76.62 (19.15)	SPIN: 31.23 (9.85) LSAS-SR: 50.38 (17.60)	SPIN: -1.091 LSAS-SR: -1.370	No-intervention control *(12)*	SPIN: 37.50 (8.01) LSAS-SR: 72.25 (20.33)	SPIN: 40.00 (11.39) LSAS-SR: 67.83 (28.38)	SPIN: 0.312 LSAS-SR: -0.217

Romano (2020)	33	Imagery rescripting *(11)* Imaginal exposure *(11)*	-	-	-	Supportive counseling *(11)*	-	-	-
Rubin (2009)	39	Imagery exposure; Positively enhanced *(10)* Negatively supplemented dosed *(10)* Prolonged *(10)* Dosed *(9)*	trial 1, after iteration 1: positively enhanced dosed exposure (PDE): SUDs: 5.0 Facial EMG: 6.58 negatively supplemented dosed exposure (NDE): SUDs: 6.9 Facial EMG: 7.22 prolonged exposure (PE): SUDs: 6.4 Facial EMG: 7.22; dosed exposure (DE): SUDs: 5.3 Facial EMG: 7.31	trial 10, after iteration 10: positively enhanced dosed exposure (PDE): SUDs: 3.0 Facial EMG: 7.18 negatively supplemented dosed exposure (NDE): SUDs: 5.0 Facial EMG: 7.77 prolonged exposure (PE): SUDs: 5.1	PDE -	-	-	-	-

				Facial EMG: 8.70 dosed exposure (DE): SUDs: 3.1 Facial EMG: 7.33					
Schwartz (1987)	52	Imagery exposure (unclear)	PRCS: 23.3 (4.5) S-R Inventory: 46.9 (6.8) Speech thermometer: 5.6 (1.7) TBCL: 54.9 (14.6) Speech HR: 96.0 (14.8) Speech SCR: 3.5 (2.9)	Imaginal exposure combined with exercise PRCS: 21.2 (4.4) S-R Inventory: 40.5 (7.7) Speech thermometer: -1.6 (1.8) TBCL: 53.2 (8.8) Speech HR: 21.4 (9.4)	IE + Exercise: PRCS: -0.4667 S-R Inventory: -0.941 Speech thermometer: -4.235 Speech HR: -5.041 Speech SCR: -0.655 IE: PRCS: -0.444 S-R Inventory: -0.765 Speech	Exercise condition: neutral imagery (unclear) Placebo control: neutral imagery (unclear)	PRCS: 23.3 (4.5) S-R Inventory: 46.9 (6.8) Speech thermometer: 5.6 (1.7) TBCL: 54.9 (14.6) Speech HR: 96.0 (14.8) Speech SCR: 3.5 (2.9)	Exercise control group PRCS: 22.1 (4.4) S-R Inventory: 42.3 (7.7) Speech thermometer: 0.0 (2.3) TBCL: 51.2 (8.8) Speech HR: 9.4 (5.0) Speech SCR: 3.8 (3.6) Placebo control PRCS: 19.6 (4.4.) S-R Inventory: 42.5 (7.6) Speech thermometer:	Exercise control group PRCS: -0.267 S-R Inventory: -0.677 Speech thermometer: -3.294 TBCL: -0.253 Speech HR: -5.851 Speech SCR: 0.103

				Speech SCR: 1.6 (3.6) Imaginal exposure PRCS: 21.3 (4.4.) S-R Inventory: 41.7 (7.6) Speech thermometer: -0.8 (1.4) TBCL: 45.8 (8.8) Speech HR: 19.2 (9.5) Speech SCR: 4.0 (3.6)	thermometer: -3.765 TBCL: -0.623 Speech HR: -5.189 Speech SCR: 0.172			0.3 (1.0) TBCL: 61.8 (8.8) Speech HR: 8.2 (5.1) Speech SCR: 3.1 (3.6)	
Takanashi 2020)	25	Imagery rescripting *(25)*	SFNE: median and range = 47 (37-58)	SFNE: median = 45 (29-57)	-	-	-	-	-
Vrielynck (2009)	49	Unique/episodic imaginary exposure *(25)*	Unique/episodic: SUDS: 24.56 (20.96)	Unique/episodic:	-	-	-	-	-

		Generic imaginary exposure *(24)*	Generic: SUDS: 17.38 (11.04)	SUDS: unclear Generic: SUDS: unclear				
Weissberg (1977)	62	Desensitization *(10)* Desensitization with coping imagery *(10)*	Desensitization: BCL: 80.40 (13.73) PRCS: 22.30 (4.34) ACL: 14.50 (2.91) Desensitization with coping imagery: BCL: 62.10 (17.27) PRCS: 22.20 (2.60) ACL: 15.10 (2.30)	Desensitization: BCL: 70.80 (11.36) PRCS: 15.80 (6.29) ACL: 8.70 (3.26) Desensitization with coping imagery: BCL: 54.10 (8.09) PRCS: 16.40 (6.05) ACL: 8.90 (3.73)	Desensitization: BCL: -.699 PRCS: -1.498 ACL: -1.993 Desensitization with coping imagery: BLC: -.463 PRCS: -2.231 ACL: -2.696	Cognitive modification *(10)* Vicarious desensitization *(10)* Vicarious desensitization with imagery *(11)* Vicarious cognitive modification *(12)* No treatment control (unclear)	No treatment control group: BCL: unclear PRCS: unclear ACL: unclear Cognitive modification: BCL: 66.60 (7.47) PRCS: 21.40 (4.01) ACL: 15.40 (2.91) Vicariously desensitization: BCL: 75.20 (9.14) PRCS: 22.10 (3.51) ACL: 14.40 (3.53) Vicariously desensitization with coping imagery: BCL: 74.91 (11.18) PRCS: 21.36 (4.23)	No treatment control group: BCL: 66.70 (13.73) PRCS: 19.80 (3.34) ACL: 13.30 (2.37) Cognitive modification: BCL: 58.00 (9.88) PRCS: 12.70 (5.97) ACL: 8.10 (2.39) Vicariously desensitization: BCL: 67.70 (4.61) PRCS: 18.40 (4.32) ACL: 9.50 (2.94) Vicariously desensitization with coping imagery: BCL: 60.45 (10.31)	No treatment control: BCL: unclear PRCS: unclear ACL: unclear Cognitive modification: BCL: -1.151 PRCS: -2.170 ACL: -2.509 Vicariously desensitization: BCL: -0.821 PRCS: -1.054 ACL: -1.388 Vicariously desensitization With coping imagery:

							ACL: 15.36 (3.65) Vicariously cognitive modification: BCL: 71.08 (8.89) PRCS: 24.00 (3.72) ACL: 16.58 (2.36)	PRCS: 15.27 (6.81) ACL: 10.00 (4.95) Vicariously cognitive modification: BCL: 60.42 (7.43) PRCS: 16.92 (7.61) ACL: 10.33 (3.09)	BCL: -1.293 PRCS: -1.440 ACL: -1.468 Vicariously cognitive modification: BCL: -1.199 PRCS: -1.903 ACL: -2.649
Wild (2007)	14	Imagery rescripting *(14)*	Social phobia weekly summary scale: 5.02 (1.16) Social cognitions frequency: 58.38 (11.87) Social cognitions total belief score: 815.63 (320.04)	follow up (n = 8) Social phobia weekly summary scale: 4.02 (1.46) Social cognitions frequency: 47.50 (9.29) Social cognitions total	Social phobia: -0.862 Social cognitions: -0.917 Social cognitions/belief total: -0.656	-	-	-	-

				belief score: 605.63 (275.26)					
Wild (2008)	11	Imagery rescripting *(11)*	LSAS-CA anxiety: 64.09 (23.96) FNE: 23.91 (5.56)	post-rescripting session LSAS-CA anxiety 37.05 (29.45) FNE: - follow-up after rescripting session LSAS-CA anxiety: 40.45 (27.0) FNE: 17.91 (10.26)	LSAS-CA: -1.129 (post-session) LSAS-CA: 0.265 (after session follow-up) FNE: -	-	-	-	-

Abbreviations: not applicable (-), eye movement (EM), eye movement desensitization and reprocessing (EMDR), waitlist condition (WC), imaginal exposure (IE), previous to (pre-), posterior to (post-), subjective units of discomfort (SUDS), Fear of Negative Evaluation questionnaire (FNE), Brief Fear of Negative Evaluation questionnaire (BFNE), Social Interaction Anxiety Scale (SIAS), Liebowitz Social Anxiety Scale (LSAS), Liebowitz Social Anxiety Scale for Children and Adolescents (LSAS-CA), Affect Adjective Checklist (ACL), Personal Report of Confidence as a Speaker (PRCS), the Behavior Checklist (BCL), Child Behavior Checklist (TBCL), Stimulus-Response Inventory (S-R Inventory), Social Phobia Inventory (SPIN), facial electromyography (Facial EMG),

Schedule for Affective Disorders and Schizophrenia for School-Aged Children (K-SADS), positively enhanced dosed exposure (PDE), Social Interaction Phobia Scale (SIPS),

Table 4

Between-group outcomes

Study characteristics			Social anxiety outcomes			Imagery outcomes		
Author (year)	Study Design	*N (total)*	Effect size (Cohen's d) Intervention group	Effect size (Cohen's d) Control group	$d_{control}$-$d_{intervention}$	Effect size (Cohen's d) Intervention group	Effect size (Cohen's d) Control group	$d_{intervention}$-$d_{ncontrol}$
Engelhard (2012)	Within-subjects experimental design	29	-	-	-	Vividness: 0.173 Emotionality: -0.532	Vividness: 0.233 Emotionality: -0.149	Vividness: 0.060 Emotionality: 0.383
Homer (2018)	RCT	27	-2.560	-2.485	0.075	Vividness: - Emotionality: EM: -0.441 No EM: -0.124	-	-
Hyett (2018)	RCT	58	SIPS: -1.369	SIPS: -0.624	0.745	-	-	-

Knutsson (2019)	RCT	27	BFNE-II: -0.438 LSAS-SR: -0.391	BFNE-II: -1.069 LSAS-SR: -0.473 Anxiety: -3.028	BFNE-II: -0.631 LSAS-SR: -0.046 Anxiety: unclear	-	-	-
Lee (2013)	RCT	23	K-SADS: -1.262 K-BFNE: -1.788 LSAS-Fear: -0.576 LSAS-Avoidance: -1.165	K-SADS: -8.251 K-BFNE: -0.291 LSAS-Fear: -2.816 LSAS-Avoidance: -0.161	K-SADS: -.6989 K-BFNE: 1.497 LSAS-Fear: -2.240 LSAS-Avoidance: 1.004	Vividness: -1.953 Emotionality: -6.061	Vividness: -0.990 Emotionality: -2.168	Vividness: 0.963 Emotionality: 3.893
Nilsson (2012)	RCT	14	FNE: -1.542 SIAS: -0.436	FNE: 0.114 SIAS: 0.023	FNE: 1.656 SIAS: 0.459	Vividness: -0.556 Emotionality: -2.021	Vividness: 0.064 Emotionality: -0.519	Vividness: 0.620 Emotionality:1.502
Norton (2016)	RCT	60	SIAS: -0.720 BFNE: -0.422	Cognitive restructuring control SIAS: -1.098 B-FNE: -1.316 Control SIAS: -0.294 B-FNE: -0.675	SIAS: -0.378 BFNE: -0.894	Vividness: -0.235 Emotionality: -0.471	CR: Vividness: -0.086 CR: Emotionality: 0.015 Control Vividness: -0.011 Emotionality: -0.050	Intervention VS CR: Vividness: 0.149 Emotionality: 0.486 Intervention VS Control: Vividness: 0.224 Emotionality: 0.421
Norton (2021)	Case series	15	SIAS + SPS: -0.031 BFNE: -0.350	SIAS + SPS: -0.737 BFNE: -0.353	SIAS + SPS: 0.706 BFNE: 0.003	Vividness: - Emotionality/impact of self-image: Positive: 0.817	-	-

						Negative: -0.913 Intensity: 0.233 Embarrassed/ashamed: -1.629 Pleased/proud: 3.629		
Reimer (2015)	RCT	25	SPIN: -1.091 LSAS-SR: -1.370	SPIN: 0.312 LSAS-SR: -0.217	SPIN: 1.403 LSAS-SR: 1.153	Vividness: - Emotional impact: positive: 2.455 negative: -2.116 intense: 0.587 unclear	Vividness: - Emotional impact positive: 0 negative: -0.797 intense: -0.433 unclear	Vividness: - Emotional impact Positive: -.2455 Negative: 1.319 Intense: -1.02 Unclear
Schwartz (1987)	RCT	52	IE + Exercise: PRCS: -0.467 S-R Inventory: -0.941 Speech thermometer: -4.235 Speech HR: -5.041 Speech SCR: -0.655 IE: PRCS: -0.444 S-R Inventory: -0.765 Speech thermometer: -	Exercise control group PRCS: -0.267 S-R Inventory: -0.677 Speech thermometer: -3.294 TBCL: -0.253 Speech HR: -5.851 Speech SCR: 0.103	IE + Exercise VS control: PRCS: 0.200 S-R Inventory: -0.264 Speech thermometer: 0.940 Speech HR: -0.810 IE VS control: PRCS: 0.177 S-R Inventory: -0.088 Speech thermometer: 7.059			

			3.765 TBCL: -0.623 Speech HR: -5.189 Speech SCR: 0.172		TBCL: 0.370 Speech HR: -0.662			
Weissberg (1977)	RCT	62	Desensitization: BCL: -0.699 PRCS: -1.498 ACL: -1.993 Desensitization with coping imagery: BLC: -0.463 PRCS: -2.231 ACL: -2.696	No treatment control: BCL: unclear PRCS: unclear ACL: unclear Cognitive modification: BCL: -1.151 PRCS: -2.170 ACL: -2.509 Vicariously desensitization: BCL: -0.821 PRCS: -1.054 ACL: -1.388 Vicariously desensitization	Desensitization VS no treatment control: BCL: unclear PRCS: unclear ACL: unclear Desensitization VS cognitive modification: BLC: -0.452 PRCS: -0.672 ACL: -0.516 Desensitization VS vicariously desensitization: BLC: 1.520 PRCS: 0.444 ACL: 0.605 Desensitization VS	-	-	-

With coping imagery: BCL: -1.293 PRCS: -1.440 ACL: -1.468 Vicariously cognitive modification: BCL: -1.199 PRCS: -1.903 ACL: -2.649	vicariously desensitization With coping imagery: BLC: -0.594 PRCS: 0.058 ACL: 0.525 Desensitization vs vicariously cognitive modification: BLC: -0.500 PRCS: -0.405 ACL: -0.656 Desensitization with coping imagery VS no treatment control: BCL: unclear PRCS: unclear ACL: unclear Intervention VS cognitive modification: BLC: -0.688 PRCS: -0.061

ACL: 0.817

Desensitization with coping imagery VS vicariously desensitization:

BLC: 1.284

PRCS: 1.177

ACL: 1.308

Desensitization with coping imagery VS vicariously desensitization

With coping imagery:

BLC: -0.830

PRCS: 0.791

ACL: 1.228

Desensitization with coping imagery vs vicariously cognitive modification:

BLC: -0.736

PRCS: 0.328

					ACL: 0.047			
Wild (2008)	Case series	11	LSAS-CA: -1.129 (post-session) LSAS-CA: 0.265 (after session follow-up) FNE: -	LSAS-CA: -0.043	LSAS-CA: 1.086	-	Memory distress: -0.281	Unclear

Abbreviations: not applicable (-), eye movement (EM), eye movement desensitization and reprocessing (EMDR), waitlist condition (WC), imaginal exposure (IE), previous to (pre-), posterior to (post-), auditory (Aud.), subjective units of discomfort (SUDS), Fear of Negative Evaluation questionnaire (FNE), Brief Fear of Negative Evaluation questionnaire (BFNE), Social Interaction Anxiety Scale (SIAS), Liebowitz Social Anxiety Scale (LSAS), Liebowitz Social Anxiety Scale for Children and Adolescents (LSAS-CA), Affect Adjective Checklist (ACL), Personal Report of Confidence as a Speaker (PRCS), the Behavior Checklist (BCL), Child Behavior Checklist (TBCL), Stimulus-Response Inventory (S-R Inventory), Social Phobia Inventory (SPIN), facial electromyography (Facial EMG), Schedule for Affective Disorders and Schizophrenia for School-Aged Children (K-SADS), positively enhanced dosed exposure (PDE), Social Interaction Phobia Scale (SIPS),

Methodological quality assessment

The methodological quality of the included studies (N = 21) was assessed using quality assessment forms for Randomized Controlled Trials (Appendix 2) and for Quasi-Experiments (Appendix 3). The studies were randomized controlled trials (n =15) and quasi-experiments or case series (n = 6). The quality assessment analysis can be found in Appendix 4. From all RCT's, two studies didn't use true randomization off participants and one study was unclear in weather or not they used true randomization. Reviewers found that for the sixteen RCT's, those delivering treatment were only blind to treatment assignment in a few cases (n = 4). This was mostly due to computerized tasks, as opposed to therapist-guided interventions where the therapist couldn't be blind to treatment. In the found RCT's, some studies showed that their outcome assessors were blind tot treatment assignment (n = 4). For the RCT's, treatment groups were treated identically other than the intervention of interest in more than half of the studies (n = 10). Reviewers stated that outcomes were measured in reliable ways in most RCT's (n = 13) and appropriate statistical analysis was used in more than half of the RCT's (n = 9). Those that were found to not have used appropriate statistical analysis were mostly because of low power/power problems (n = 5). Of all quasi-experiments (n = 6), the cause and effect of the studies was clear in most quasi-experiments (n = 5). None of the quasi-experiments had a control group or control period, and only one quasi-experiment had done a follow-up. Half of quasi-experiments did not use appropriate statistical analysis due to low power (n = 3).

Discussion

This thesis aimed to review the characteristics of imagery interventions and their effects on (sub-)clinical social anxiety and imagery characteristics. To assess the research question, 'What are the characteristics and outcomes of imagery-based interventions in social anxiety?', a systematic review was executed and the results included twenty-one studies and

their extracted data. The data provided an overview of information about which interventions were used, how images were targeted within the interventions and also provided information about the effect of the intervention on image emotionality and vividness as well as social anxiety measurements.

Primary findings

When looking at the characteristics of imagery-based interventions, the interventions that were most found in the included studies were based on imagery rescripting (n = 12), desensitization (n = 1), eye movement desensitization and reprocessing (n = 5) and imaginal exposure (n = 3) (see Table 1). Most imagery interventions were therapist-guided (n = 16) whilst some were computerized tasks (n = 5). The types of images targeted in the intervention were autobiographical memories (*n* = 14), script driven images (*n* = 3), a problematic recent social anxiety event (*n* = 1), a flashforward image regarding social anxiety (*n* = 1) a public speaking situation (*n* = 1) and an intrusive social anxiety image (*n* = 1). From all twenty-one included studies, fourteen studies included either a control group (*n* =10) and/or a control period (*n* = 4). Eight studies had no control group (*n* = 8). This gives quite a good overview of the content of the images that were targeted with the imagery interventions. Imagery rescripting was the most used intervention within the included studies. Most intervention targeted an autobiographical memory that they had linked to their current social anxiety imagery. However, sometimes a script-driven memory was used to induce an image. The duration of the intervention varied from a few minutes to multiple hours. Though imagery rescripting and EMDR were the most used interventions, their duration differed. The duration of the imagery rescripting interventions ranged from 30-45 min 9 (*n* = 2), to 60-90min (*n* = 1) or 90+min (*n* = 7). For EMDR interventions, the duration ranged from 60 seconds (*n* = 2) to 2-3 minutes (*n* = 2) to 45min (*n* = 1) to 90 min (*n* = 1). Control conditions existed of no-treatment control (*n* = 2), verbal restructuring (*n* = 1), eye movements with neutral imagery

without desensitization and reprocessing (n = 3), group cognitive behavioural therapy (n = 1), cognitive modification (n = 1), in vivo exposure (n = 1), waitlist control (n = 1), exercise condition (n = 1), supportive counselling (n = 1), a control period of 3 weeks (n = 1) and vicarious video treatments (desensitization, n = 1, desensitization with coping imagery, n = 1, and cognitive modification, n = 1). Research states that waitlist control is the most-used control group used in published psychological treatment studies, compared to attention placebo or the best available treatment comparison (Paterson, Boyle, Kivienieks & van Ameringen, 2016). Patterson and colleagues also state that effect sizes are much higher in trials when compared to waitlists than to psychological placebos (Patterson et al., 2016).

When looking at the outcomes of imagery-based interventions for social anxiety, this review included a variety of outcome measurements. Studies with clinical samples mostly used clinical questionnaires to assess social anxiety, while studies with subclinical samples used non-clinical questionnaires for social anxiety symptoms.

The results showed medium to large effect sizes for both imagery characteristics (emotionality and vividness) and social anxiety measurements. There was a reduction in both social anxiety and imagery characteristics, as almost all of the reported effect sizes were negative. It seems that imagery rescripting interventions often had a longer duration per intervention than other interventions. EMDR-interventions sometimes seemed to have a small duration of only a couple minutes. Another thing the imagery rescripting studies have in common is a longer duration of the intervention (total) time in most studies.

As for social anxiety measures, most studies showed negative intervention effect sizes indicating a reduction of either social anxiety or image emotionality or vividness, suggesting that these interventions targeting negative imagery appear to have an effect on social anxiety symptoms, measured in various ways. However, there were also some positive

effect sizes, indicating an increase in measured aspects. Namely in the study of Homer et al., (2016) in almost all intervention conditions; one medium positive effect size, four small positive effect sizes for the measurements of social anxiety. In Wild et al., (2008), even though the post-session effect size for social anxiety was negative and large (d = -1.129), indicating a large reduction of social anxiety, the follow-up measurement of the intervention group was a small positive effect of d = 0.265. It seems as if the effect of the intervention on the reduction of social anxiety decreased with time. This shows that follow-up measurements can differ from post-intervention measurements, and that the time-effect of interventions should be taken into account more often to provide more information about how long the intervention will have effect on symptoms.

As for the imagery measurements, the information provided by the included studies about the targeted images is limited as well as what the intervention goal was related to the images. Some studies did however report how the image was gained. Some were trough an interview which could be spontaneous or with a more structured interview to target the image. Other studies tracked memories of social anxiety situations and connected an image with the memory, other provided their participants with a scripted negative mental image related to social anxiety. With script-driven interventions, it is not always clear whether or not the participants have influence on the chosen image, for instance if they do not recognize the scripted situation as fearful. Some studies do provide a more detailed content of their used intervention, such as Kearns & Engelhard, (2015), Carrigan & Levis, (1999), Nilsson et al., (2012), Rubin et al., (2009) and Takanashi et al., (2020). However, other studies provide little information about the content of their intervention, such as Frets et al., (2014) and Homer et al., (2016). For the use of imagery-interventions in practice, a more detailed protocol of how the image is perceived is seen as more useful as it is easier to reproduce in further research.

This leads to the belief that studies who do not provide information about how the image was targeted, should be evaluated with care. Their effect sizes say something, to some extent, about the effect of the imagery-intervention on their outcome measures. However, due to missing information about the origin and content of the images, it will be more difficult to say if the nature of the image (recent memory, childhood memory, aversive social event, script-driven image) has any influence on the effect of the intervention itself. The number of dropouts per intervention and control condition was also included in the data extraction. However, most studies did not report dropouts, thus the number of dropouts was unclear in most studies. This data was not included in the data extraction tables because it provided very little information. Interestingly, could imply a publication bias due to the number of dropouts being unclear so often.

Limitations

A statistic limitation is the number of unclear values in the included studies. These missing values are sometimes due to studies not measuring both social anxiety and imagery characteristics, sometimes due to missing values such as standard deviations which makes calculating an effect size more difficult. Furthermore, seven studies did not include a control period or control group, making it more difficult to state the intervention effect because confounding variables are not controlled. This was also seen in the quality assessment. This missing data makes it more difficult to make statements about the overall effect sizes of all the included studies. Interestingly, the study of Weissberg (1977) used five different control groups, showing a negative (mostly large) effect size in almost all five control groups as well. However, the control groups of Weissberg (1997) were all some form of psychological therapy/aid, except for one no-treatment control group, which lacked information to calculate an effect size due to missing pre-intervention measurements of social anxiety.

Another statistic limitation is that there was a plan to conduct a meta-analysis of the studies as well. However, because it ended up being beyond the scope of this study, no meta-analysis was conducted. When possible, it was planned to conduct a meta-analysis using studies with a control group only. This meta-analysis, if possible, would be conducted using Jeffreys's Amazing Statistics Program, known as JASP (JASP Team, 2020). Heterogeneity would be assessed as well. When there was at least mild heterogeneity found, a subgroup analysis would be conducted. This could be done using the following possible subgroups: 1) clinical or subclinical samples, 2) ages below 18 or above 18, 3) types of interventions and 4) types of targeted images. A meta-analysis could be conducted when at least two studies with a control group report similar pre- and post-intervention measurements of social anxiety and/or similar pre- and post-interventions measurements of image characteristics such as vividness, amount of emotionality or other imagery characteristics. Together with this limitation, it was also planned to analyse the references of included articles to possibly find more articles that could fit the inclusion criteria. Within the scope of this research, it was no longer possible to check the references of the included articles. For future research, it is advised to do so if the scope and time allow so.

Another limitation is the inclusion of clinical and subclinical social anxiety as criteria for studies. By using subclinical social anxiety as inclusion criteria, it may be more difficult to generalize the outcomes of this review to clinical populations. It is therefore also more difficult to state whether or not social anxiety as a DSM-5 diagnose was de- or increased by the imagery interventions in the studies. However, this inclusion criteria did provide a broad set of articles and possibly made participants more willing to participate in such studies, seen as clinical populations has a more ethical issue by including said population in a trial intervention. Also, by including a variety of imagery interventions instead of choosing one

imagery intervention as the main focus, less can be said per imagery intervention. If, for instance, the review was focused on only imagery rescripting only, more could have been said about the effect of imagery rescripting on social anxiety. Because imagery rescripting intervention-studies appeared to provide the largest overall effect sizes, it is recommended to follow up imagery rescripting as an intervention for social anxiety to more narrowly look into which characteristics of imagery rescripting provide such effect sizes. However, it has been useful to contain a broad view within this paper as to be able to determine which intervention shows which effect sizes as opposed to other interventions. By including all imagery interventions that met the inclusion criteria as they were in this review, it provided a useful comparison of effect sizes between different interventions.

Another limitation is the quality assessment. Even though most studies got a 'Yes' to most quality assessment questions, the research would have been more valuable if the quality assessment also used as a cut off, saying that a study must at least fulfill a certain amount of quality assessment questions to be included in as a qualitative study. Now, the quality assessment was used more to analyze the studies than to select them. However, this method of assessing quality might also give a more complete overview of the studies that were included and how, for instance, studies with a lower quality can also provide information about the intervention. For example, a case study often gives a more detailed explanation of the intervention per case, however, if using a cut off, a case study probably would not have been part of the included studies. Van Heeren, Mogoase, Philippot & McNally (2015) did, for example, pair a value of unclear, low or high risk of bias to their included and assessed articles. Though this seems informative, labeling a study with a high risk of bias might also be a quick conclusion as the quality of a study sometimes goes beyond the scope of a quality checklist. For example, the study of Frets et al. (2014), was more difficult to rate in the

quality assessment. However, they did provide detailed information about their patients and their characteristics, as well as about their imagery intervention. Frets et al. (2014) used graphs to portrait their pre- and post-interventions measurements of social anxiety for their participants. They did not provide exact scores, making it more difficult to interpret the results and thereby the quality assessment question whether or not the researchers used appropriate statistical analysis. The graphs show a reduction of the measured social anxiety, however, due to exact numbers missing, this reduction is difficult to interpret.

A methodological limitation within this thesis is that within the scope of this research, a meta-analysis was not conducted. For future research, a meta-analysis using the data conducted in this review could provide more information about possible subgroups within the research such as age, gender, intervention type, duration of intervention and used measurement of social anxiety and imagery characteristics.

Interestingly, most studies did not meet the assessor-blindness within their study. However, this seems logically explained due to the nature of psychological interventions. One needs training to perform certain interventions, therefor it is expected that those delivering treatment are not blind to the content of the treatment. For this same reason, participant blindness is also difficult, especially when studies involve a waitlist-control or no-treatment control group or period. Also, three out of four quasi-experiments did not involve pre- and post-measurements of the primary outcomes in the intervention group or did not provide pre- and post-intervention measurements of the control group. The within group effect sizes were still calculated and reported, however, the between-group measurements were reported in difference scores because it was outside the scope of this thesis to report overall controlled effect sizes of all included studies. This means that within this thesis, also no analysis was made on heterogeneity or subgroup effects, as stated in the method.

Conclusion

In the current systematic review, imagery rescripting was found to be effective on both social anxiety measures and imagery characteristics vividness and emotionality. Interestingly, emotionality often showed larger effect sizes than vividness. Medium and even large effects were found for various interventions targeting negative imagery in individuals with social anxiety. Even though almost all effect sizes were negative, indicating a reduction of social anxiety as an effect of imagery interventions, more information is needed about the influence of duration, content and targeted image to provide a more practical overview of the effective imagery interventions on social anxiety individuals.

Implications for clinical practice and research

Taken together, more research is required to determine which characteristics of imagery interventions have an effect in reducing (sub-)clinical social anxiety. Specifically, more research is needed about the content and frequency of negative mental imagery in social anxiety as well as why imagery rescripting seems to be so effective. There is a lot of variation in duration of intervention, it would be useful to know at what duration an intervention has the largest long-term effect on reducing social anxiety. To conduct more information about this, more randomized controlled trials are needed in this area to test the applicability of imagery interventions in practice. Within randomized controlled trials, it is important that researchers report their pre- and post-intervention measurements clearly and are explicit in what the effect of the intervention is on social anxiety as well as on imagery vividness and/emotionality. Because there were larger effects found in reducing imagery emotionality than image vividness, more research is needed as to why this is and what effect this has on symptoms, perhaps reducing emotionality alone is also a good predictor of reducing social

anxiety. For this research goal, it is also needed that researchers report clearly on their outcomes.

As for clinical implications, this review shows a that in most studies, imagery interventions have an effect on social anxiety. Because imagery-interventions are not yet standard-treatment within social anxiety, more research is needed to include image-based interventions in clinical practice, as is could be beneficial in reducing social anxiety symptoms and image emotionality. Much research has already been done about imagery in anxiety disorders in for instance depression and PTSD (Hirsch & Holmes, 2007). Imagery rescripting is a widely known effective treatment for unprocessed traumatic images, often connected to traumatic memories, for PTSD (Hagenaars & Arntz, 2012). Early (inadequate) memory processing plays a large role in PTSD development (Brewin, 2001). Though PTSD and social anxiety have negative mental imagery as well as negatively labeling memories in common, the treatment in social anxiety is not yet focused on treating that negative mental imagery. Both mental health professionals and their clients would benefit from more attention for negative mental imagery integrated in social anxiety treatment. In an ideal world, targeting images would be implicated in standard social anxiety treatment in mental health.

References

Bisson, J.I., Ehlers, A. Mathews, A., Pilling, S., Richards, D., Turner, S. (2007). Psychological treatments for chronic post-traumatic stress disorder: systematic review and meta-analysis. *British Journal of Psychiatry*, *190*, 97-104

Brewin, Chris. (2001). A cognitive neuroscience model of posttraumatic stress disorder and its treatment. Behaviour research and therapy. 39. 373-93. 10.1016/S0005-7967(00)00087-5.

Carrigan, M. H., & Levis, D. J. (1999). The contributions of eye movements to the efficacy of brief exposure treatment for reducing fear of public speaking. *Journal of Anxiety Disorders*, *13*(1-2), 101-118.

Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg, M. R. Liebowitz, D. A. Hope, & F. R. Schneier (Eds.), Social phobia: Diagnosis, assessment, and treatment (pp. 69 – 93). New York, NY7 Guilford Press.

Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org

Engelhard, I. M., Sijbrandij, M., Van Den Hout, M. A., Rutherford, N. M., Rahim, H. F., & Kocak, F. (2012). Choking under pressure: Degrading flashforwards related to performance anxiety. *Journal of Experimental Psychopathology*, *3*(2), 158-167.

Fehm, L., Pelissolo, A., Furmark, T., & Wittchen, H. U. (2005). Size and burden of social phobia in Europe. *European neuropsychopharmacology*, *15*(4), 453-462.

Frets, P. G., Kevenaar, C. & van der Heiden, C. (2014). Imagery rescripting as a stand-alone treatment for patients with social phobia: A case series. *Journal of Behavior Therapy and Experimental Psychiatry, 45*(1), 160-169. doi: 10.1016/j.jbtep.20013.09.006.

Hackmann, A., Surawy, C., & Clark, D. M. (1998). Seeing yourself through other's eyes: a study of spontaneously occurring images in social phobia. *Behavioural and Cognitive Psychotherapy*, *26*(1), 3-12.

Hagenaars, M. A., & Arntz, A. (2012). Reduced intrusion development after post-trauma imagery rescripting; an experimental study. *Journal of behavior therapy and experimental psychiatry*, *43*(2), 808-814.

Heeren, A., Mogoașe, C., Philippot, P., & McNally, R. J. (2015). Attention bias modification for social anxiety: a systematic review and meta-analysis. *Clinical psychology review*, *40*, 76-90.

Hirsch, C. R., & Holmes, E. A. (2007). Mental imagery in anxiety disorders. *Psychiatry*, *6*(4), 161-165.

Holmes, E. A. & Mathews, A. (2005). Mental Imagery and Emotion: A Special Relationship? *Emotion*, 5(4), 489-497.

Homer, S. R., & Deeprose, C. (2018). Eye movement attenuation of intrusive social anxiety imagery: A pilot study.

Homer, S. R., Deeprose, C., & Andrade, J. (2016). Negative mental imagery in public speaking anxiety: Forming cognitive resistance by taxing visuospatial working memory. *Journal of behavior therapy and experimental psychiatry*, *50*, 77-82.

Hyett, M. P., Bank, S. R., Lipp, O. V., Erceg-Hurn, D. M., Alvares, G. A., Maclaine, E., ... & McEvoy, P. M. (2018). Attenuated psychophysiological reactivity following single-session group imagery rescripting versus verbal restructuring in social anxiety disorder:

results from a randomized controlled trial. *Psychotherapy and psychosomatics*, *87*(6), 340-349.

JASP Team (2020). JASP (Version 0.14.1) [Computer software].

Joanna Briggs Institute, (2020. *Checklist for Randomized Controlled Trials.* https://jbi.global/critical-appraisal-tools

Joanna Briggs Institute, (2020). *Checklist for Quasi-Experimental Studies*. https://jbi.global/critical-appraisal-tools

Kearns, M., & Engelhard, I. M. (2015). Psychophysiological responsivity to script-driven imagery: An exploratory study of the effects of eye movements on public speaking flashforwards. *Frontiers in psychiatry*, *6*, 115.

Knutsson, J., Nilsson, J. E., Eriksson, Å., & Järild, L. (2019). Imagery rescripting and exposure in social anxiety: a randomized trial comparing treatment techniques. *Journal of contemporary psychotherapy*, 1-8.

Lee, S. W., & Kwon, J. H. (2013). The efficacy of imagery rescripting (IR) for social phobia: a randomized controlled trial. *Journal of behavior therapy and experimental psychiatry*, *44*(4), 351-360.

Nilsson, J. E., Lundh, L. G., & Viborg, G. (2012). Imagery rescripting of early memories in social anxiety disorder: An experimental study. *Behaviour Research and Therapy*, *50*(6), 387-392.

Norton, A. R., & Abbott, M. J. (2016). The efficacy of imagery rescripting compared to cognitive restructuring for social anxiety disorder. *Journal of anxiety disorders*, *40*, 18-28.

Norton, A. R., Abbott, M. J., Dobinson, K. A., Pepper, K. L., & Guastella, A. J. (2021). Rescripting Social Trauma: A Pilot Study Investigating Imagery Rescripting as an

Adjunct to Cognitive Behaviour Therapy for Social Anxiety Disorder. *Cognitive Therapy and Research*, 1-13.

Ohanian, V. (2001). Imagery rescripting within cognitive behavior therapy for bulimia nervosa: An illustrative case report. *International Journal of Eating Disorders*, *31*, 352-357.

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

Page, M.J., McKenzie J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D, (2020). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136

Patterson, B., Boyle, M. H., Kivlenieks, M., & Van Ameringen, M. (2016). The use of waitlists as control conditions in anxiety disorders research. *Journal of psychiatric research*, *83*, 112-120.

Pittig, A., Wong, A. H., Glück, V. M., & Boschet, J. M. (2020). Avoidance and its bi-directional relationship with conditioned fear: Mechanisms, moderators, and clinical implications. *Behaviour research and therapy*, *126*, 103550.

Reimer, S. G., & Moscovitch, D. A. (2015). The impact of imagery rescripting on memory appraisals and core beliefs in social anxiety disorder. *Behaviour Research and Therapy*, *75*, 48-59.

Romano, M., Moscovitch, D. A., Huppert, J. D., Reimer, S. G., & Moscovitch, M. (2020). The effects of imagery rescripting on memory outcomes in social anxiety disorder. *Journal of anxiety disorders*, *69*, 102169.

Rubin, S., Spates, C. R., Johnson, D. A., & Jouppi, L. (2009). Dosed versus prolonged exposure in the treatment of fear: An experimental evaluation and review of behavioral mechanisms. *Journal of Anxiety Disorders*, *23*(6), 806-812.

Sagaltici, E., & Okan Demirci, O. (2019). Eye Movement Desensitization and Reprocessing (EMDR) Treatment in Social Anxiety Disorder: A Case Report. *Psychiatria Danubina*, *31*(3), 358-359.

Schwartz, S. G., & Kaloupek, D. G. (1987). Acute exercise combined with imaginal exposure as a technique for anxiety reduction. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *19*(2), 151.

Stein, M. B., & Stein, D. J. (2008). Social anxiety disorder. *The lancet*, *371*(9618), 1115-1125.

Takanashi, R., Yoshinaga, N., Oshiro, K., Matsuki, S., Tanaka, M., Ibuki, H., ... & Shimizu, E. (2020). Patients' perspectives on imagery rescripting for aversive memories in social anxiety disorder. *Behavioural and Cognitive Psychotherapy*, *48*(2), 229-242.

Vrielynck, N., & Philippot, P. (2009). Regulating emotion during imaginal exposure to social anxiety: Impact of the specificity of information processing. *Journal of Behavior Therapy and expérimental psychiatry*, *40*(2), 274-282.

Weissberg, M. (1977). A comparison of direct and vicarious treatments of speech anxiety: Desensitization, desensitization with coping imagery, and cognitive modification. *Behavior Therapy*, *8*(4), 606-620.

Whiting, P., Rutjes, A.W.S., Reitsma, J.B., Bossuyt, P.M.M., Kleijnen, J., (2003). The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology. doi:10.1186/1471-2288-3-25.

Wild, J., Hackmann, A., & Clark, D. M. (2007). When the present visits the past: Updating traumatic memories in social phobia. *Journal of behavior therapy and experimental psychiatry*, *38*(4), 386-401.

Wild, J., Hackmann, A., & Clark, D. M. (2008). Rescripting early memories linked to negative images in social phobia: a pilot study. Behavior Therapy, 39(1), 47e56.

Appendix

Appendix 1: Prospero Preregistration

Systematic review

Please complete all mandatory fields below (marked with an asterisk *) and as many of the non-mandatory fields as you can then click *Submit* to submit your registration. You don't need to complete everything in one go, this record will appear in your *My PROSPERO* section of the web site and you can continue to edit it until you are ready to submit. Click *Show help* below or click on the icon to see guidance on completing each section.

*** Review title.**

Interventions targeting negative mental imagery in social anxiety: a systematic review and meta-analysis of characteristics and outcomes

Original language title.

Interventies gericht op negatieve mentale beelden bij sociale angst: een systematische review en meta-analyse van kenmerken en uitkomsten

* **Anticipated or actual start date.**

09 July 2021

Anticipated completion date.

11 January 2022

* Stage of review at time of this submission.

The review has not yet started: Yes

Review stage	Started	Completed
Preliminary searches	**Yes**	**No**
Piloting of the study selection process	**No**	**No**
Formal screening of search results against eligibility criteria	**No**	**No**
Data extraction	**No**	**No**
Risk of bias (quality) assessment	**No**	**No**
Data analysis	**No**	**No**

Provide any other relevant information about the stage of the review here (e.g. Funded proposal, protocol not yet finalised).

-

*** Named contact.**

Ms Marjolein Thunnissen, MSc

*** Named contact email.**

m.r.thunnissen@rug.nl

Named contact address

Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, the Netherlands.

Named contact phone number.

-

Organisational affiliation of the review.

University of Groningen

Review team members and their organisational affiliations.

Marjolein Thunnissen, MSc, Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, the Netherlands. m.r.thunnissen@rug.nl

Alcinda Costa, BSc, Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, the Netherlands. a.b.costa@student.rug.nl

Prof. dr. Maaike Nauta, Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, the Netherlands. m.h.nauta@rug.nl

Prof. dr. Peter de Jong, Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, the Netherlands. p.j.de.jong@rug.nl

Prof. dr. Marleen Rijkeboer, Department of Clinical Psychological Science, University of Maastricht, the Netherlands. marleen.rijkeboer@maastrichtuniversity.nl

Dr. Marisol Voncken, Department of Clinical Psychological Science, University of

Maastricht, the Netherlands. m.voncken@maastricthuniversity.nl

Funding sources/sponsors.

This work was, in part, supported by a research fund awarded to the Marjolein Thunnissen and Maaike Nauta by the educational institution Post MSc Psychology & Orthopedagogics (Opleidingsinstituut PPO) linked to the University of Groningen. Opleidingsinstituut PPO has no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest.

None known

Collaborators.

-

Review question.

Negative mental imagery frequently occurs in individuals with social anxiety, and seems to play an important role in the persistence of social anxiety. Therefore, it would be valuable to evaluate the possibilities and effects of imagery-based interventions focused on targeting negative mental images in social anxiety. The current review and meta-analysis focuses on evaluating the different interventions that have been applied to target negative imagery in social anxiety and on evaluating the outcomes with regards to social anxiety and imagery characteristics, aiming to answer the following research question:

What are the characteristics and outcomes of interventions targeting negative mental imagery in social anxiety?

To the best of our knowledge, no previous reviews have been conducted with this scope.

Searches.

Electronic searches will be conducted in PsycINFO and MEDLINE databases. The planned search date is 9 July 2021. Language will be restricted to English. No restrictions will be set for publication year. We plan to use combinations of the following search terms:

"social anxiety disorder" or "social phobi*" or "social* anxi*" or "SAD" or "performance anxiety" or "speech anxiety" or "speaking anxiety"

and

"negative imag*" or "imagery-based" or "imag* intervention" or "imag* training" or "imag* therapy" or "imag* exposure" or "imagery rescripting" or "ImRs" or "imag* morphing" or "emotive imagery" or "EMDR" or "eye movement desensitization therapy" or "eye movement desensitization and reprocessing" or "dual-task*" or "competing task"

OF

"social anxiety disorder" or "social phobi*" or "social* anxi*" or "SAD" or "performance anxiety" or "speech anxiety" or "speaking anxiety"

And

"imagery" or "mental imag*" or "negative imag*" or "negative memor*" or "imagery-based"

And

"intervention" or "training" or "therapy" or "imag* exposure" or "imagery rescripting"

or “ImRs” or “imag* morphing” or “emotive imagery” or "EMDR" or "eye movement desensitization therapy" or "eye movement desensitization and reprocessing" or “dual-task*” or “competing task”

The references of included articles resulting from the searches will also be checked for relevant titles. We will not search for unpublished studies.

URL to search strategy.

-

Condition or domain being studied.

Clinical or subclinical social anxiety and negative mental imagery. Social anxiety includes performance anxiety types.

Participants/population.

Inclusion: The study concerns individuals with clinical or subclinical social anxiety; clinical social anxiety includes an established social anxiety disorder diagnosis, for example, using a clinical interview according to the DSM criteria; subclinical social anxiety includes elevated scores on a measure of social anxiety symptoms, for example, using a self-report questionnaire. Social anxiety includes performance anxiety types. No restrictions are made with regards to age or gender.

Exclusion: The study does not concern social anxiety; e.g., the study concerns individuals with another primary diagnosis than social anxiety disorder, or the study concerns subclinical symptoms of another condition than social anxiety.

Intervention(s), exposure(s).

Inclusion: The study concerns an intervention targeting negative mental images in social anxiety, defined as an intervention that is imagery-based and focuses on reducing (the impact of) relevant negative mental imagery in social anxiety. Examples are forms of imagery rescripting and EMDR with relevant negative mental imagery (including relevant negative memories) as a target. No further restrictions are made with regards to the type, form, and duration of an intervention.

Exclusion: The study does not concern an intervention with a primary focus on targeting negative imagery; e.g., no more than 50% of the intervention consists of targeting negative imagery using an imagery-based approach, or the intervention focuses on positive or other imagery, or the study focuses on inducing rather than reducing (the impact of) negative mental imagery.

Comparator(s)/control.

The study either includes a control group (e.g., waitlist, treatment as usual, active intervention) as control, a baseline control period or no control (e.g., pre- and post-design).

Types of study to be included.

Inclusion: Primary, empirical studies published in an academic journal that include quantitative pre- and post-intervention assessments of social anxiety and/or characteristics of negative mental imagery.

Exclusion: Studies including only qualitative assessments, studies without empirical data such as reviews or study protocols, dissertations, and non-English studies.

Context.

-

*** Main outcome(s).**

Change in social anxiety from pre-intervention to post-intervention. Change in imagery characteristics from pre-intervention to post-intervention. Establishing characteristics of interventions targeting negative imagery in social anxiety.

* Measures of effect

Change in main outcomes (social anxiety and imagery characteristics) from pre- to post-intervention time points

*** Additional outcome(s).**

None

* Measures of effect

Not applicable

*** Data extraction (selection and coding).**

Studies resulting from the searches will be screened for eligibility using the above-mentioned in- and exclusion criteria in the following steps. The first step consists of screening based on titles and abstracts, which will be conducted independently by two reviewers from the review team. Disagreements will be discussed and resolved by reaching consensus with the help of a third reviewer of the review team. Studies that are selected based on this first step, will continue to the second step of screening based on the full-text. The full-text will also be screened independently by two reviewers from the review team for eligibility, and disagreements will also be discussed and resolved by reaching consensus with the help of a third reviewer.

The data from selected eligible studies will be extracted independently by two independent reviewers from the review team. Disagreements will be resolved by consensus. The data that will be extracted from each study includes:

- Study characteristics, study design (including control group, control period or no control), number of participants (also for the control group, when applicable), sample characteristics (including clinical or subclinical group and average age > or <18), social anxiety measure(s), imagery measure(s)

- Intervention characteristics (e.g., type, form, imagery target(s), amount)

- Intervention effects: Mean, standard deviation, and effect size of social anxiety pre- and post-intervention time points (also for the control group or period, when applicable), mean, standard deviation, and effect size of imagery vividness and emotionality pre- and post-intervention time points (also for the control group or period, when applicable), mean, standard deviation, and effect size of other imagery characteristics pre- and post-intervention time points (also for the control group or period, when applicable), dropout from the intervention (and from the control group, when applicable)

When multiple measures are used to assess pre- and post-intervention social anxiety, the effects data of the social anxiety symptom measures that are stated as main outcome measures in the study for social anxiety is extracted. The same principles are applied for imagery vividness, imagery emotionality, and other imagery characteristics.

The study selection and data extraction process will be conducted in Covidence and guided by the PRISMA guidelines.

*** Risk of bias (quality) assessment.**

Risk of bias for each study will be assessed independently by two reviewers of the team, and disagreements will be resolved by consensus. The JBI Critical Appraisal Tools (including different tools for RCT's and quasi-experimental studies) will be used. This will also be conducted in Covidence.

*** Strategy for data synthesis.**

When possible, a meta-analysis will be conducted using only studies with a control group. Effect sizes will be calculated for pre-intervention to post-intervention changes in social anxiety and/or imagery characteristics compared to the control group.

A meta-analysis is conducted when at least two studies with a control group report similar measurements of social anxiety pre- and post-intervention and/or similar measurements of imagery vividness, imagery emotionality or other imagery characteristics pre- and post-intervention.

The outcomes measured in that case will be the effect sizes of the pre-post changes in mean social anxiety and/or imagery characteristics compared to the control group.

We will calculate pooled effect sizes using a random effects model.

Heterogeneity will be assessed. In case of at least mild heterogeneity, we will conduct planned subgroup analyses. Furthermore, we will assess publication bias.

*** Analysis of subgroups or subsets.**

Planned subgroup analyses will be conducted in case of at least mild heterogeneity and at least two studies within the subgroups. Planned subgroups include: studies with clinical and subclinical groups, age >18 or <18, intervention type, and intervention imagery target.

*** Type and method of review.**

Type of review

Cost effectiveness	No
Diagnostic	No
Epidemiologic	No
Individual patient data (IPD) meta-analysis	No
Intervention	Yes
Meta-analysis	Yes

Methodology	No
Narrative synthesis	No
Network meta-analysis	No
Pre-clinical	No
Prevention	No
Prognostic	No
Prospective meta-analysis (PMA)	No
Review of reviews	No
Service delivery	No
Synthesis of qualitative studies	No

Systematic review	Yes
Other	No

Health area of the review

Alcohol/substance misuse/abuse	No
Blood and immune system	No
Cancer	No
Cardiovascular	No
Care of the elderly	No
Child health	No
Complementary therapies	No
Crime and justice	No

Dental	No
Digestive system	No
Ear, nose and throat	No
Education	No
Endocrine and metabolic disorders	No
Eye disorders	No
General interest	No
Genetics	No
Health inequalities/health equity	No
Infections and infestations	No

International development	No
Mental health and behavioral conditions	Yes
Musculoskeletal	No
Neurological	No
Nursing	No
Obstetrics and gynecology	No
Oral health	No
Palliative care	No
Perioperative care	No
Physiotherapy	No

Pregnancy and childbirth	No		
Public health (including social determinants of health)	No	Rehabilitation	No
Respiratory disorders	No		
Service delivery	No		
Skin disorders	No		
Social care	No		
Surgery	No		
Tropical Medicine	No		
Urological	No		
Wounds, injuries and accidents	No		

Violence and abuse	No

Language.

English

*** Country.**

The Netherlands

Other registration details.

-

Reference and/or URL for published protocol.

-

Dissemination plans.

The review is intented to be published. / Yes

Keywords.

Social anxiety; social phobia; imagery interventions; imagery rescripting; EMDR; mental images; negative imagery; systematic review; meta-analysis.

Details of any existing review of the same topic by the same authors.

-

*** Current review status.**

Review: Ongoing / planned

Any additional information.

-

Details of final report/publication(s).

-

Appendix 2: Quality assessment: Critical appraisal checklists

Checklist for Randomized Controlled Trials

The Checklist for Randomized Controlled Trials was used to assess included studies that were RCTs (Joanna Briggs Institute, 2020).

JBI CRITICAL APPRAISAL CHECKLIST FOR RANDOMIZED CONTROLLED TRIALS

Reviewer_____
Date_____

Author_____Year_____ Record Number_____

	Yes	No	Unclear	NA
1. Was true randomization used for assignment of participants to treatment groups?	☐	☐	☐	☐
2. Was allocation to treatment groups concealed?	☐	☐	☐	☐
3. Were treatment groups similar at the baseline?	☐	☐	☐	☐
4. Were participants blind to treatment assignment?	☐	☐	☐	☐
5. Were those delivering treatment blind to treatment assignment?	☐	☐	☐	☐
6. Were outcomes assessors blind to treatment assignment?	☐	☐	☐	☐
7. Were treatment groups treated identically other than the intervention of interest?	☐	☐	☐	☐
8. Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed?	☐	☐	☐	☐
9. Were participants analyzed in the groups to which they were randomized?	☐	☐	☐	☐
10. Were outcomes measured in the same way for treatment groups?	☐	☐	☐	☐
11. Were outcomes measured in a reliable way?	☐	☐	☐	☐
12. Was appropriate statistical analysis used?	☐	☐	☐	☐

13. Was the trial design appropriate, and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial? ☐ ☐ ☐ ☐

Overall appraisal: Include ☐ Exclude ☐ Seek further info ☐

Comments (Including reason for exclusion)

_____

_____

_____

_____

Explanation for the critical appraisal tool for RCTs with individual participants in parallel groups

Answers: Yes, No, Unclear or Not/Applicable

Critical Appraisal Tool for RCTs (individual participants in parallel groups)

Was true randomization used for assignment of participants to treatment groups?

The differences between participants included in compared groups constitutes a threat to the internal validity of a study exploring causal relationships. If participants are not allocated to treatment and control groups by random assignment there is a risk that the allocation is influenced by the known characteristics of the participants and these differences between the groups may distort the comparability of the groups. A true random assignment of participants to the groups means that a procedure is used that allocates the participants to groups purely based on chance, not influenced by the known characteristics of the participants. Check the details about the randomization procedure used for allocation of the participants to study groups. Was a true chance (random) procedure used? For example, was a list of random numbers used? Was a computer-generated list of random numbers used?

Was allocation to groups concealed?

If those allocating participants to the compared groups are aware of which group is next in the allocation process, that is, treatment or control, there is a risk that they may deliberately and purposefully intervene in the allocation of patients by preferentially allocating patients to the treatment group or to the control group and therefore this may distort the implementation of allocation process indicated by the randomization and therefore the results of the study may be distorted. Concealment of allocation (allocation concealment) refers to procedures that prevent those allocating patients from knowing before allocation which treatment or control is next in the allocation process. Check the details about the procedure used for allocation concealment. Was an appropriate allocation concealment procedure used? For example, was central randomization used? Were sequentially numbered, opaque and sealed envelopes used? Were coded drug packs used?

Were treatment groups similar at the baseline?

The differences between participants included in compared groups constitute a threat to the internal validity of a study exploring causal relationships. If there are differences between participants included in compared groups there is a risk of selection bias. If there are differences between participants included in the compared groups maybe the ‘effect’ cannot

be attributed to the potential ‘cause’ (the examined intervention or treatment), as maybe it is plausible that the ‘effect’ may be explained by the differences between participants, that

is, by selection bias. Check the characteristics reported for participants. Are the participants from the compared groups similar with regards to the characteristics that may explain the effect even in the absence of the 'cause', for example, age, severity of the disease, stage of the disease, co-existing conditions and so on? Check the proportions of participants with specific relevant characteristics in the compared groups. Check the means of relevant measurements in the compared groups (pain scores; anxiety scores; etc.). *[Note: Do NOT only consider the P-value for the statistical testing of the differences between groups with regards to the baseline characteristics.]*

Were participants blind to treatment assignment?

If participants are aware of their allocation to the treatment group or to the control group there is the risk that they may behave differently and respond or react differently to the intervention of interest or to the control intervention respectively compared to the situations when they are not aware of treatment allocation and therefore the results of the study may be distorted. Blinding of participants is used in order to minimize this risk. Blinding of the participants refers to procedures that prevent participants from knowing which group they are allocated. If blinding of participants is used, participants are not aware if they are in the group receiving the treatment of interest or if they are in any other group receiving the control interventions. Check the details reported in the article about the blinding of participants with regards to treatment assignment. Was an appropriate blinding procedure used? For example, were identical capsules or syringes used? Were identical devices used? Be aware of different terms used, blinding is sometimes also called masking.

Were those delivering treatment blind to treatment assignment?

If those delivering treatment are aware of participants' allocation to the treatment group or to the control group there is the risk that they may behave differently with the participants from the treatment group and the participants from the control group, or that they may treat them differently, compared to the situations when they are not aware of treatment allocation and this may influence the implementation of the compared treatments and the results of the study may be distorted. Blinding of those delivering treatment is used in order to minimize this risk. Blinding of those delivering treatment refers to procedures that prevent those delivering treatment from knowing which group they are treating, that is those delivering treatment are not aware if they are treating the group receiving the treatment of interest or if they are treating any other group receiving the control interventions. Check the details reported in the article about the blinding of those delivering treatment with regards to treatment assignment. Is there any information in the article about those delivering the treatment? Were those delivering the treatment unaware of the assignments of participants to the compared groups?

Were outcomes assessors blind to treatment assignment?

If those assessing the outcomes are aware of participants' allocation to the treatment group or to the control group there is the risk that they may behave differently with the participants from the treatment group and the participants from the control group compared to the situations when they are not aware of treatment allocation and therefore there is the risk that the measurement of the outcomes may be distorted and the results of the study may be distorted. Blinding of outcomes assessors is used in order to minimize this risk. Check the details reported in the article about the blinding of outcomes assessors with regards to treatment assignment. Is there any information in the article about outcomes assessors? Were those assessing the treatment's effects on outcomes unaware of the assignments of participants to the compared groups?

Were treatment groups treated identically other than the intervention of interest?

In order to attribute the 'effect' to the 'cause' (the treatment or intervention of interest), assuming that there is no selection bias, there should be no other difference between the groups in terms of treatment or care received, other than the manipulated 'cause' (the treatment or intervention controlled by the researchers). If there are other exposures or treatments occurring at the same time with the 'cause' (the treatment or intervention of interest), other than the 'cause', then potentially the 'effect' cannot be attributed to the examined 'cause' (the investigated treatment), as it is plausible that the 'effect' may be explained by other exposures or treatments occurring at the same time with the 'cause' (the treatment of interest). Check the reported exposures or interventions received by the compared groups. Are there other exposures or treatments occurring at the same time with the 'cause'? Is it plausible that the 'effect' may be explained by other exposures or treatments occurring at the same time with the 'cause'? Is it clear that there is no other difference between the groups in terms of treatment or care received, other than the treatment or intervention of interest?

Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed?

For this question, follow up refers to the time period from the moment of random allocation (random assignment or randomization) to compared groups to the end time of the trial. This critical appraisal question asks if there is complete knowledge (measurements, observations etc.) for the entire duration of the trial as previously defined (that is, from the moment of random allocation to the end time of the trial), for all randomly allocated participants. If there is incomplete follow up, that is incomplete knowledge about all randomly allocated participants, this is known in the methodological literature as the post-assignment attrition. As RCTs are not perfect, there is almost always post-assignment attrition, and the focus of this question is on the appropriate exploration of post-assignment attrition (description of loss to follow up, description of the reasons for loss to follow up, the estimation of the impact of loss

to follow up on the effects etc.). If there are differences with regards to the loss to follow up between the compared groups in an RCT, these differences represent a threat to the internal validity of a randomized experimental study exploring causal effects, as these differences may provide a plausible alternative explanation for the observed 'effect' even in the absence of the 'cause' (the treatment or intervention of interest). When appraising an RCT, check if there were differences with regards to the loss to follow up between the compared groups. If follow up was incomplete (that is, there is incomplete information on all participants), examine the reported details about the strategies used in order to address incomplete follow up, such as descriptions of loss to follow up (absolute numbers; proportions; reasons for loss to follow up) and impact analyses (the analyses of the impact of loss to follow up on results). Was there a description of the incomplete follow up (number of participants and the specific reasons for loss to follow up)? It is important to note that with regards to loss to follow up, it is not enough to know the number of participants and the proportions of participants with incomplete data; the reasons for loss to follow up are essential in the analysis of risk of bias; even if the numbers and proportions of participants with incomplete data are similar or identical in compared groups, if the patterns of reasons for loss to follow up are different (for example, side effects caused by the intervention of interest, lost contact etc.), these may impose a risk of bias if not appropriately explored and considered in the analysis. If there are differences between groups with regards to the loss to follow up (numbers/proportions and reasons), was there an analysis of patterns of loss to follow up? If there are differences between the groups with regards to the loss to follow up, was there an analysis of the impact of the loss to follow up on the results? [Note: Question 8 is NOT about intention-to-treat (ITT) analysis; question 9 is about ITT analysis.]

Were participants analyzed in the groups to which they were randomized?

This question is about the intention-to-treat (ITT) analysis. There are different statistical analysis strategies available for the analysis of data from randomized controlled trials, such as intention-to-treat analysis (known also as intent to treat; abbreviated, ITT), per-protocol analysis, and as-treated analysis. In the ITT analysis the participants are analyzed in the groups to which they were randomized, regardless of whether they actually participated or not in those groups for the entire duration of the trial, received the experimental intervention or control intervention as planned or whether they were compliant or not with the planned experimental intervention or control intervention. The ITT analysis compares the outcomes for participants from the initial groups created by the initial random allocation of participants to those groups. Check if ITT was reported; check the details of the ITT. Were participants analyzed in the groups to which they were initially randomized, regardless of whether they actually participated in those groups, and regardless of whether they actually received the planned interventions? *[Note: The ITT analysis is a type of statistical analysis recommended in the Consolidated Standards of Reporting Trials (CONSORT) statement on best practices in trials reporting, and it is considered a marker of good methodological quality of the analysis of results of a randomized trial. The ITT is*

estimating the effect of offering the intervention, that is, the effect of instructing the participants to use or take the intervention; the ITT it is not estimating the effect of actually receiving the intervention of interest.]

Were outcomes measured in the same way for treatment groups?

If the outcome (the 'effect') is not measured in the same way in the compared groups there is a threat to the internal validity of a study exploring a causal relationship as the differences in outcome measurements may be confused with an effect of the treatment (the 'cause'). Check if the outcomes were measured in the same way. Same instrument or scale used? Same measurement timing? Same measurement procedures and instructions?

Were outcomes measured in a reliable way?

Unreliability of outcome measurements is one threat that weakens the validity of inferences about the statistical relationship between the 'cause' and the 'effect' estimated in a study exploring causal effects. Unreliability of outcome measurements is one of the different plausible explanations for errors of statistical inference with regards to the existence and the magnitude of the effect determined by the treatment ('cause'). Check the details about the reliability of measurement such as the number of raters, training of raters, the intra-rater reliability, and the inter-raters reliability within the study (not as reported in external sources). This question is about the reliability of the measurement performed in the study, it is not about the validity of the measurement instruments/scales used in the study. *[Note: Two other important threats that weaken the validity of inferences about the statistical relationship between the 'cause' and the 'effect' are low statistical power and the violation of the assumptions of statistical tests. These other two threats are explored within Question 12).]*

Was appropriate statistical analysis used?

Inappropriate statistical analysis may cause errors of statistical inference with regards to the existence and the magnitude of the effect determined by the treatment ('cause'). Low statistical power and the violation of the assumptions of statistical tests are two important threats that weaken the validity of inferences about the statistical relationship between the 'cause' and the 'effect'. Check the following aspects: if the assumptions of statistical tests were respected; if appropriate statistical power analysis was performed; if appropriate effect sizes were used; if appropriate statistical procedures or methods were used given the number and type of dependent and independent variables, the number of study groups, the nature of the relationship between the groups (independent or dependent groups), and the objectives of statistical analysis (association between variables; prediction; survival analysis etc.).

Was the trial design appropriate for the topic, and any deviations from the standard RCT design accounted for in the conduct and analysis?

Certain RCT designs, such as the crossover RCT, should only be conducted when appropriate. Alternative designs may also present additional risks of bias if not accounted for in the design and analysis.

Crossover trials should only be conducted in people with a chronic, stable condition, where the intervention produces a short-term effect (i.e. relief in symptoms). Crossover trials should ensure there is an appropriate period of washout between treatments.

Cluster RCTs randomize groups of individuals, forming ‘clusters.’ When we are assessing outcomes on an individual level in cluster trials, there are unit-of-analysis issues, as individuals within a cluster are correlated. This should be taken into account by the study authors when conducting analysis, and ideally authors will report the intra-cluster correlation coefficient.

Stepped-wedge RCTs may be appropriate when it is expected the intervention will do more good than harm, or due to logistical, practical or financial considerations in the roll out of a new treatment/intervention. Data analysis in these trials should be conducted appropriately, taking into account the effects of time.

Appendix 3: Quality Assessment Critical Appraisal Checklist for Quasi-Experimental Studies

The Checklist for Quasi-Experiments was used to assess included studies that were quasi-experiments (Joanna Briggs Institute, 2020).

JBI CRITICAL APPRAISAL CHECKLIST FOR QUASI-EXPERIMENTAL STUDIES

Reviewer_____
Date_____

Author_____Year_____ Record Number_____

		Yes	No	Unclear	Not applicable
1.	Is it clear in the study what is the 'cause' and what is the 'effect' (i.e. there is no confusion about which variable comes first)?	☐	☐	☐	☐
2.	Were the participants included in any comparisons similar?	☐	☐	☐	☐
3.	Were the participants included in any comparisons receiving similar treatment/care, other than the exposure or intervention of interest?	☐	☐	☐	☐
4.	Was there a control group?	☐	☐	☐	☐
5.	Were there multiple measurements of the outcome both pre and post the intervention/exposure?	☐	☐	☐	☐
6.	Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed?	☐	☐	☐	☐
7.	Were the outcomes of participants included in any comparisons measured in the same way?	☐	☐	☐	☐

8. Were outcomes measured in a reliable way? ☐ ☐ ☐ ☐

9. Was appropriate statistical analysis used? ☐ ☐ ☐ ☐

Overall appraisal: Include ☐ Exclude ☐ Seek further info ☐

Comments (Including reason for exclusion)

_____

_____

_____

_____

Explanation for the critical appraisal tool for quasi-experimental studies

Critical Appraisal Tool for Quasi-Experimental Studies (Experimental Studies without random allocation)

Answers: Yes, No, Unclear or Not/Applicable

1. Is it clear in the study what is the 'cause' and what is the 'effect' (i.e. there is no confusion about which variable comes first)?

Ambiguity with regards to the temporal relationship of variables constitutes a threat to the internal validity of a study exploring causal relationships. The 'cause' (the independent variable, that is, the treatment or intervention of interest) should occur in time before the explored 'effect' (the dependent variable, which is the effect or outcome of interest). Check if it is clear which variable is manipulated as a potential cause. Check if it is clear which variable is measured as the effect of the potential cause. Is it clear that the 'cause' was manipulated before the occurrence of the 'effect'?

2. Were the participants included in any comparisons similar?

The differences between participants included in compared groups constitute a threat to the internal validity of a study exploring causal relationships. If there are differences between participants included in compared groups there is a risk of selection bias. If there are differences between participants included in the compared groups maybe the 'effect' cannot be attributed to the potential 'cause', as maybe it is plausible that the 'effect' may be explained by the differences between participants, that is, by selection bias. Check the characteristics reported for participants. Are the participants from the compared groups similar with regards to the characteristics that may explain the effect even in the absence of the 'cause', for example, age, severity of the disease, stage of the disease, co-existing conditions and so on? *[NOTE: In one single group pre-test/post-test studies where the patients are the same (the same one group) in any pre-post comparisons, the answer to this question should be 'yes.']*

3. Were the participants included in any comparisons receiving similar treatment/care, other than the exposure or intervention of interest?

In order to attribute the 'effect' to the 'cause' (the exposure or intervention of interest), assuming that there is no selection bias, there should be no other difference between the

groups in terms of treatments or care received, other than the manipulated 'cause' (the intervention of interest). If there are other exposures or treatments occurring in the same time with the 'cause', other than the intervention of interest, then potentially the 'effect' cannot be attributed to the intervention of interest, as it is plausible that the 'effect' may be explained by other exposures or treatments, other than the intervention of interest, occurring in the same time with the intervention of interest. Check the reported exposures or interventions received by the compared groups. Are there other exposures or treatments occurring in the same time with the intervention of interest? Is it plausible that the 'effect' may be explained by other exposures or treatments occurring in the same time with the intervention of interest?

4. Was there a control group?

Control groups offer the conditions to explore what would have happened with groups exposed to other different treatments, other than to the potential 'cause' (the intervention of interest). The comparison of the treated group (the group exposed to the examined 'cause', that is, the group receiving the intervention of interest) with such other groups strengthens the examination of the causal plausibility. The validity of causal inferences is strengthened in studies with at least one independent control group compared to studies without an independent control group. Check if there are independent, separate groups, used as control groups in the study. *[Note: The control group should be an independent, separate control group, not the pre-test group in a single group pre-test post-test design.]*

5. Were there multiple measurements of the outcome both pre and post the intervention/exposure?

In order to show that there is a change in the outcome (the 'effect') as a result of the intervention/treatment (the 'cause') it is necessary to compare the results of measurement before and after the intervention/treatment. If there is no measurement before the treatment and only measurement after the treatment is available it is not known if there is a change after the treatment compared to before the treatment. If multiple measurements are collected before the intervention/treatment is implemented then it is possible to explore the plausibility of alternative explanations other than the proposed 'cause' (the intervention of interest) for the observed 'effect', such as the naturally occurring changes in the absence of the 'cause', and changes of high (or low) scores towards less extreme values even in the absence of the 'cause' (sometimes called regression to the mean). If multiple measurements are collected after the intervention/treatment is implemented it is possible to explore the changes of the 'effect' in

time in each group and to compare these changes across the groups. Check if measurements were collected before the intervention of interest was implemented. Were there multiple pre-test measurements? Check if measurements were collected after the intervention of interest was implemented. Were there multiple post-test measurements?

6. Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed?

If there are differences with regards to the loss to follow up between the compared groups these differences represent a threat to the internal validity of a study exploring causal effects as these differences may provide a plausible alternative explanation for the observed 'effect' even in the absence of the 'cause' (the treatment or exposure of interest). Check if there were differences with regards to the loss to follow up between the compared groups. If follow up was incomplete (that is, there is incomplete information on all participants), examine the reported details about the strategies used in order to address incomplete follow up, such as descriptions of loss to follow up (absolute numbers; proportions; reasons for loss to follow up; patterns of loss to follow up) and impact analyses (the analyses of the impact of loss to follow up on results). Was there a description of the incomplete follow up (number of participants and the specific reasons for loss to follow up)? If there are differences between groups with regards to the loss to follow up, was there an analysis of patterns of loss to follow up? If there are differences between the groups with regards to the loss to follow up, was there an analysis of the impact of the loss to follow up on the results?

7. Were the outcomes of participants included in any comparisons measured in the same way?

If the outcome (the 'effect') is not measured in the same way in the compared groups there is a threat to the internal validity of a study exploring a causal relationship as the differences in outcome measurements may be confused with an effect of the treatment or intervention of interest (the 'cause'). Check if the outcomes were measured in the same way. Same instrument or scale used? Same measurement timing? Same measurement procedures and instructions?

8. Were outcomes measured in a reliable way?

Unreliability of outcome measurements is one threat that weakens the validity of inferences about the statistical relationship between the 'cause' and the 'effect' estimated in a study

exploring causal effects. Unreliability of outcome measurements is one of different plausible explanations for errors of statistical inference with regards to the existence and the magnitude of the effect determined by the treatment

('cause'). Check the details about the reliability of measurement such as the number of raters, training of raters, the intra-rater reliability, and the inter-raters reliability within the study (not to external sources). This question is about the reliability of the measurement performed in the study, it is not about the validity of the measurement instruments/scales used in the study. *[Note: Two other important threats that weaken the validity of inferences about the statistical relationship between the 'cause' and the 'effect' are low statistical power and the violation of the assumptions of statistical tests. These other threats are not explored within Question 8, these are explored within Question 9.]*

9. Was appropriate statistical analysis used?

Inappropriate statistical analysis may cause errors of statistical inference with regards to the existence and the magnitude of the effect determined by the treatment ('cause'). Low statistical power and the violation of the assumptions of statistical tests are two important threats that weakens the validity of inferences about the statistical relationship between the 'cause' and the 'effect'. Check the following aspects: if the assumptions of statistical tests were respected; if appropriate statistical power analysis was performed; if appropriate effect sizes were used; if appropriate statistical procedures or methods were used given the number and type of dependent and independent variables, the number of study groups, the nature of the relationship between the groups (independent or dependent groups), and the objectives of statistical analysis (association between variables; prediction; survival analysis

Appendix 3: Quality Assessment Table

Author (year)	Quality assessment questions												
Carrigan (1999)	RCT 1: Yes	RCT2: Unclear	RCT3: Yes	RCT4: Yes	RCT5: No	RCT: No	RCT7: Yes	RCT8: No	RCT9: No	RCT10: Yes	RCT11: Yes	RCT12: Yes	RCT13: Yes
Engelhard (2012)	QE1: Yes	QE2: Yes	QE3: Yes	QE4: No	QE5: No	QE6: NA	QE7: Yes	QE8: Yes	QE9: No	NA	NA	NA	NA
Frets (2014)	QE1: Yes	QE2: No	QE3: No	QE4: No	QE5: Yes	QE6: Unclear	QE7: Yes	QE8: Unclear	QE9: Unclear	QE10: NA	NA	NA	NA
Homer (2016)	RCT1: Yes	RCT2: Yes	RCT3: Yes	RCT4: Yes	RCT5: Yes	RCT6: Yes	RCT7: Yes	RCT8: Unclear	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: Yes	RCT13: Yes
Homer (2018)	RCT1: Yes	RCT2: Yes	RCT3: Yes	RCT4: Yes	RCT5: Yes	RCT6: Unclear	RCT7: Yes	RCT8: NA	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: No	RCT13: Yes
Hyett (2018)	RCT1: Yes	RCT2: Yes	RCT3: Not applicable	RCT4: No	RCT5: No	RCT6: Yes	RCT7: Yes	RCT8: No	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: Yes	RCT13: Yes

Kearns (2015)	RCT1: Yes	RCT2: Yes	RCT3: Yes	RCT4: Yes	RCT5: Yes	RCT6: Yes	RCT7: Yes	RCT8: NA	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: Yes	RCT13: Yes
Knutsson (2019)	RCT1: No	RCT2: Yes	RCT3: Yes	RCT4: Unclear	RCT5: Unclear	RCT6: No	RCT7: Yes	RCT8: Yes	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: Yes	RCT13: Yes
Lee (2013)	RCT1: Yes	RCT2: Yes	RCT3: Yes	RCT4: Yes	RCT5: Yes	RCT6: Unclear	RCT7: No	RCT8: Yes	RCT9: Yes	QE10:	QE11:	NA	NA
Nilsson (2012)	RCT1: Yes	RCT2: Unclear	RCT3: Yes	RCT4: Yes	RCT5: Yes	RCT6: Unclear	RCT7: Yes	RCT8: Yes	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: Yes	RCT13: Yes
Norton (2016)	RCT1: Yes	RCT2: Yes	RCT3: Yes	RCT4: No	RCT5: No	RCT6: Unclear	RCT7: Yes	RCT8: Yes	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: Yes	RCT13: Yes
Norton (2021)	QE1: Yes	QE2: Yes	QE3: Yes	QE4: No	QE5: No	QE6: Unclear	QE7: Yes	QE8: Yes	QE9: Yes	NA	NA:	NA	NA
Reimer (2015)	RCT1: Yes	Unclear	RCT3: Yes	RCT4: No	RCT5: No	RCT6: Yes	RCT7: Yes	RCT8: Yes	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: No	RCT13: Yes
Romano (2020)	RCT1: Yes	RCT2: Yes	RCT3: No	RCT4: Unclear	RCT5: No	RCT6: Unclear	RCT7: Yes	RCT: Yes	RCT9: Yes	QE10:	QE11:	RCT12: Yes	RCT13: Yes
Rubin (2009)	RCT1: Yes	RCT2: Unclear	RCT3: Yes	RCT4: Yes	RCT5: Yes	RCT6: Yes	RCT7: No	RCT8: Yes	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: Yes	RCT13: Yes
Schwartz (1987)	RCT1: Unclear	RCT2: Unclear	RCT3: Yes	RCT4: Yes	RCT5: No	RCT6: Unclear	RCT7: Yes	RCT8: Unclear	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: No	RCT13: Yes

Takanashi 2020)	QE1: Yes	QE2: Yes	QE3: Unclear	QE4: No	QE5: No	QE6: Unclear	QE7: Yes	QE8: Yes	QE9: Yes	NA	NA	NA	NA
Vrielynck (2009)	RCT1: Yes	RCT2: Unclear	RCT3: Yes	RCT4: Yes	RCT5: No	RCT6: No	RCT7: Yes	RCT8: Yes	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: Yes	RCT13: Yes
Weissberg (1977)	RCT1: No	RCT2 No	RCT3: Yes	RCT4: No	RCT5: No	RCT6: Unclear	RCT7: Yes	RCT8: Unclear	RCT9: Yes	RCT10: Yes	RCT11: Yes	RCT12: No	RCT13: Unclear
Wild (2007)	QE1: Yes	QE2: Yes	QE3: Yes	QE4: No	QE5: Yes	QE6: No	QE7: Yes	QE8: Yes	QE9: Yes	NA	NA	NA	NA
Wild (2008)	QE1: Yes	QE2: Yes	QE3: Yes	QE4: No	QE5: Yes	QE6: Yes	QE7: Yes	QE8: Yes	QE9: Yes	NA	NA	NA	NA

Abbreviations: Randomized controlled trial (RCT), Quasi-experiment (QE), not applicable (NA). For the abbreviations of the quality assessment questions, see Appendix 2 and 3.