Review

Scrambled but valid? The scrambled sentences task as a measure of interpretation biases in psychopathology: A systematic review and meta-analysis

Felix Würtz [*], Lisa Zahler, Simon E. Blackwell, Jürgen Margraf, Mahdi Bagheri, Marcella L. Woud

Mental Health Research and Treatment Center, Faculty of Psychology, Ruhr-Universität Bochum, Bochum, Germany

ABSTRACT

The Scrambled Sentences Task (SST) is frequently used to assess interpretation biases (IBs). However, neither the range of its applications nor the quality of the empirical evidence it provides has been systematically examined. This systematic review investigates the types of samples and disorders in which the SST has been applied and evaluates its psychometric properties via a meta-analysis. The databases PubMed and EBSCOhost (including PsycINFO, PsycARTICLES, PSYNDEX, MEDLINE) were examined (last search: September 2021) and 93 studies from 91 manuscripts were included. Results showed that the SST has been applied predominantly in unselected samples or those with elevated levels of subsyndromal symptoms, with about a third of the studies employing the SST in a clinical population. While the SST was initially developed to assess depression-related IBs, it has now been extended to other disorders, in particular anxiety disorders. Results of the meta-analyses indicated good convergent validity and reliability across disorders, albeit in the context of substantial heterogeneity. Findings concerning divergent validity were mixed with high correlations across disorders between the SST and trait anxiety in particular, questioning its specificity. Future research should consider developing standardized SST versions and investigating its relationships with other measures of IB.

1. Introduction

Interpretation biases (IBs) play a central role in many cognitive models of psychopathology (e.g., Beck, 1976; Clark, 1986); how we measure them is therefore a crucial aspect of psychopathology research. IBs are the tendency to interpret ambiguous information in a systematically biased manner, for example consistently negatively, and are observed across many mental disorders (for reviews, see Everaert, Podina, & Koster, 2017; Hirsch, Meeten, Krahé, & Reeder, 2016; Mathews & MacLeod, 2005). Research has found that IBs not only often correlate with symptoms of psychopathology (e.g., Korn, Dietel, & Hartmann, 2020; Lee, Mathews, Shergill, & Yiend, 2016), but also predict future symptoms and even the onset of a disorder (Kleim et al., 2013; Woud, Zhang, Becker, McNally, & Margraf, 2014). Further, experimental manipulation of IBs can lead to changes in analog, and over time, actual symptoms of psychopathology (e.g., Hirsch et al., 2018; Joormann, Waugh, & Gotlib, 2015; Van Bockstaele, Clarke,

Notebaert, MacLeod, & Salemink, 2020) albeit not consistently across the literature (Fodor et al., 2020; Jones & Sharpe, 2017). In summary, for some disorders at least, there is accumulating evidence that IBs may fulfil the criteria to be considered causal factors in psychopathology (Kraemer et al., 1997). Their assessment therefore has relevance across all areas of research, from theory testing and development through to treatment.

A large number of methods has been developed to measure IBs (for an overview, see Table S1, Supplementary Material; for review, see Hirsch et al., 2016), and one that is widely used is the Scrambled Sentences Task (SST; Wenzlaff & Bates, 1998). The SST typically requires participants to sort five out of six words into a grammatically correct sentence. Depending on the omitted word, the resulting sentence represents a particular (valenced) interpretation of the original ambiguous stimulus (e.g., in the context of depression: winner a born I am loser; positive interpretation: I am a born *winner*; negative interpretation: I am a born *loser*). The SST is usually completed with a time limit and under

* Corresponding author at: Mental Health Research and Treatment Center, Faculty of Psychology, Ruhr-Universität Bochum, Massenbergstrasse 9-13, 44787 Bochum, Germany.
E-mail address: felix.wuertz@rub.de (F. Würtz).

cognitive load, e.g., retaining a 6-digit number in mind, which is intended to interfere with controlled stimulus processing and thus reduce deliberate response biases (Bowler et al., 2012; Rude, Valdez, Odom, & Ebrahimi, 2003). The SST was first developed and applied in the context of depression (Wenzlaff & Bates, 1998), and meta-analyses indicate strong associations between scores on the SST, for example the percentage of sentences completed negatively, and symptoms or diagnosis of depression (Everaert et al., 2017; Phillips, Hine, & Thorsteinsson, 2010). Beyond depression, the SST has been increasingly applied in other disorder contexts such as Generalized Anxiety Disorder (GAD; Krahé, Whyte, Bridge, Loizou, & Hirsch, 2019), psychosis (Savulich, Freeman, Shergill, & Yiend, 2015), and anxiety sensitivity (Zahler et al., 2020). In summary, the SST is a commonly applied measure of IBs that is designed to overcome some limitations of self-report measures such as response or demand effects.

While research using the SST suggests its utility and promise as a measure of IB across many different disorders, the SST has not yet been systematically examined. We therefore aimed to provide a systematic overview of the different ways in which the SST had been applied, including types of samples, disorder domains, and other characteristics (e.g., language). We also aimed to examine and evaluate the SST's psychometric properties such as its validity and reliability. To this end we conducted a systematic literature review of studies using the SST together with disorder-specific and unspecific symptom measures, across various areas of psychopathology. Further we conducted two meta-analyses, one for indices of convergent validity and one for those of internal consistency.

2. Methods

This review was conducted in accordance with the PRISMA guidelines (Liberati et al., 2009). A study protocol was uploaded prior to the literature search (retrievable from the Open Science Framework: osf.io/4fkjs). After the full-text screening, however, an amendment was made to the protocol, to also include doctoral dissertations if they met the eligibility criteria. The decision to synthesize the information extracted via a meta-analytic approach was made based on reviewer feedback during peer-review and was not part of the original protocol. All procedures deviating from the protocol are labeled as such. The OSF also includes all other relevant materials, e.g., the extracted information, the analysis scripts (excluding scripts involving unpublished data from other authors), and the data extraction template.

2.1. Eligibility criteria

Studies were considered eligible when they were written in English or German, and when the SST was used to measure IBs in combination with disorder-relevant symptom measures. There were no eligibility criteria regarding the studies' publication year. Further, studies were included independently of their current publication status (e.g., we also included data of papers in preparation) to reduce the impact of a potential publication bias.

2.2. Information sources and search strategy

The databases PubMed and EBSCOhost (including PsycINFO, PsycARTICLES, PSYNDEX, and MEDLINE) were used for the literature search, using the following search strings– *PubMed:* ((((Scrambled Sentence Task) OR Scrambled Sentences Task) OR Scrambled Sentence Test) OR Scrambled Sentences Test) OR SST; *EBSCOhost:* TX SST OR TX scrambled sentences task OR TX scrambled sentences test OR TX scrambled sentence test OR TX scrambled sentence task. Additionally, potential studies were identified through reference lists of other (eligible) publications. To include further (e.g., unpublished) studies, an e-mail using the mailing lists of the German Psychological Society (DGPs) was sent in July 2019. Further, we contacted colleagues who we

knew were working with the SST. The first search was run in July 2019. During the review's further preparation and revision, the searches were then updated four times using the same search string and starting from the timepoint of the prior search, with the last and fifth search being run on September 10th 2021.

2.3. Study selection

The selection process was as follows: Two authors independently scanned the titles and abstracts of the identified studies during the initial searches. All titles were scanned separately for EBSCOhost and PubMed after which duplicates were removed by the Mendeley duplicate detection. Remaining duplicates were removed manually. Second, the studies' eligibility was discussed and if no consensus was achieved, the third author (SEB) was consulted. Next, this procedure was repeated for the full texts. After inclusion of the full texts, two authors extracted the relevant data for each study (e.g., correlation coefficients, sample sizes, etc.) independently. Deviations between the extracted data were resolved through discussion.

For the PRISMA Flow Diagram of the selection process, see Fig. 1. Studies were counted as individual studies when they reported data of a unique sample. Thus, multiple papers reporting data of one sample were counted as one study and multiple samples reported within one paper were counted as multiple studies.

2.4. Extracted data and analytical approach meta-analysis

In line with our first aim, i.e., providing a systematic overview of how the SST has been applied so far, we extracted information related to the overall study design and the general application of the SST (e.g., language or application format). In line with our second aim, i.e., examining the SST's psychometrics via a meta-analysis, we extracted the relevant psychometric indices per study. That is, for convergent validity, we extracted correlations between SST scores and the symptom measures of the corresponding target disorder. For divergent validity, we extracted correlations between the SST and disorder-unspecific symptom measures and transdiagnostic symptoms (e.g., trait-anxiety). Importantly, these correlations were only included if (a) the SST and the other measure were assessed at the same time point (e.g., during the same laboratory or online session), (b) no manipulation (e.g., a mood induction, treatment, CBM etc.) was delivered before or between the assessment of both measures, and (c) no instruction was given as to which type of sentences participants should form (e.g., the instruction to only form positive sentences). Further, we included only zero order correlations.[1] The procedure for the extraction of the reliability indices was as follows: If a study reported reliability for multiple time points or different sets of sentences these studies were counted multiple times. When reliability indices were not reported, we contacted authors to request these indices or corresponding raw data and, when available, calculated the split-half reliability (following recommendations of Parsons, Kruijt, & Fox, 2019) using the R package 'psych' (Revelle, 2019), using the brute force method and coding grammatical errors or unfinished sentences as missing values. Next, we conducted two meta-analyses, one for the convergent validity indices and one for the reliability (internal consistency) indices. To account for multiple validity/reliability indices reported within individual studies, we conducted a multilevel-meta analysis using Maximum Likelihood estimation including random effects for within- (level 2) and between-study variability (level 3), using the *rma.mv* function of the R-package 'metafor' (Viechtbauer, 2010) and following the procedure specified by Assink and Wibbelink (2016). For convergent validity, we used correlations between the SST and symptom measures of the respective target

---

[1] Please note that these restrictions were not part of the pre-registered protocol but were decided on after the study selection procedure.
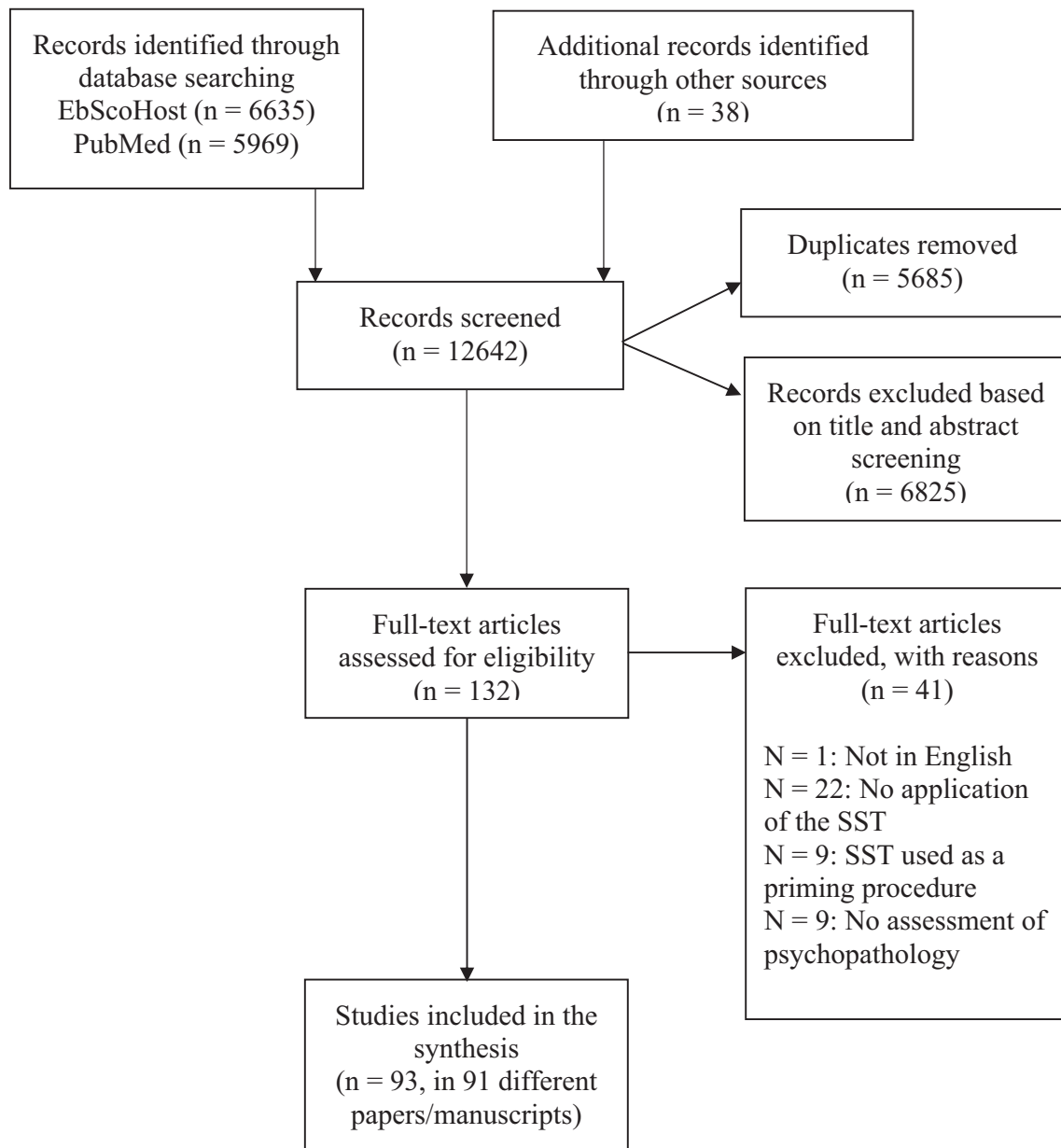
Fig. 1. PRISMA flow diagram.

disorder, and for internal consistency, we used Cronbach's alpha and split-half reliabilities. We did not conduct a meta-analysis on divergent validity since the included studies used rather heterogeneous sets of questionnaires to assess symptoms not directly related to the target disorder. As a consequence of this, the results of a meta-analysis would be very difficult to interpret.

All data was extracted separately for different disorder domains. When IBs of different domains were targeted via the SST in a single study (i.e., trials targeting worry or depression, or paranoia and general emotional content in the same study) and separate correlations/reliability indices were reported, this study was included in each suitable disorder-domain. Therefore, when grouping the studies by disorder, the sum of studies and participants exceeds the corresponding total number.

2.5. Risk of bias assessment

The risk of bias (ROB) of individual studies was assessed using a modified version of the study quality assessment checklist for cross-sectional studies of the National Institute of Health (National Heart Lung and Blood Institute, 2021). Adaptations were made as such that items were removed that did not fit this review's focus, and additional items concerning the reporting of technical aspects of the SST were added. The ROB was assessed independently by two authors and deviations were resolved through discussion (the modified checklist is available on the OSF (osf.io/4fkjs).[2] For further details on the ROB assessment see Supplementary Material.

3. Results

3.1. Populations and samples

We included 93 individual studies from 91 published papers,

---

[2] The assessment of the ROB deviates from the procedure specified in the pre-registration as the planned procedure was identified as suboptimal during peer-review.

unpublished manuscripts, preprints, or doctoral dissertations in our review, including a total sample of $n = 9913$ participants. These studies were conducted between 1998 and 2021. The majority of studies investigated depression related IBs, with fewer than ten studies in both Generalized Anxiety Disorder (GAD) and Social Anxiety Disorder (SAD) and single studies in other disorder domains. The majority of studies investigated adult samples ($k = 87$, $n = 9414$), but there were also a few studies including children and adolescents ($k = 4$, $n = 343$) and one study ($k = 1$, $n = 156$) included children and their parents. Further, a third of the studies ($k = 30$, $n = 2159$) included clinical samples, i.e., participants who were either in- or out-patients at a mental health service or had been diagnosed with a psychological disorder assessed via a clinical interview, and these participants were then compared to a non-diagnosed, non-clinical control group. The samples of the remaining studies ($k = 63$, $n = 7754$) were not formally diagnosed (e.g., community samples or participants above a questionnaire cut-off). Finally, 56.99% of the studies ($k = 53$, $n = 5548$) used an English version of the SST[3] and 39.78% of the studies ($k = 37$, $n = 4067$) applied the SST in a fully computerized form (for a detailed overview, see Table 1).

3.2. Validity of the SST

3.2.1. Convergent validity

We extracted 95 indices representing convergent validity from 53 studies of which one index was excluded due to missing sample size information (see Table 1 for a detailed overview). Results of our meta-analysis showed that the overall correlation for convergent validity was moderate to high, r [95%-CI] = 0.46 [0.41, 0.51], $p < .001$. The overall heterogeneity was high, $Q(93) = 662.25$, $p < .001$, and the variance within studies was significant (level 2: $\sigma^2 = 0.008$, 22.58%), $\chi^2(1) = 37.28$, $p < .001$, as well as the variance between studies (level 3: $\sigma^2 = 0.022$, 64.10%), $\chi^2(1) = 19.39$, $p < .001$. Finally, 13.32% of the variance was accounted for by random sampling (level 1).[4] Given these high levels of heterogeneity, we conducted exploratory post-hoc moderation analyses to better understand whether the relationship between SST and target symptom measures was related to specific study characteristics (as recommended by Assink & Wibbelink, 2016). We found that the application format (computerized vs. other) was a significant moderator of the overall correlation, $F(1, 92) = 9.62$, $p = .003$, with higher coefficients in the fully computerized version with an estimated difference in r [95%-CI] = 0.14 [0.05, 0.24], $SE = 0.05$. However, the type of disorder (depression vs. other disorder) was not associated with differences in convergent validity, $F(1, 92) = 0.50$, $p = .482$, and neither was the language in which the SST was applied (English vs. other), $F(1, 92) = 3.50$, $p = .065$, nor the application of a cognitive load[5] (i.e., present vs. absent), $F(1, 91) = 0.24$, $p = .629$, or the type of sample (partly diagnosed participants versus non-diagnosed participants),[6] $F(1,$

92) $= 0.48$, $p = .490$. Finally, the number of trials on the SST was not associated with differences in the convergent validity, $F(1, 92) = 2.43$, $p = .122$.

Comparing the intercept only model to the full model including the application format as a moderator significantly reduced the heterogeneity, $\chi^2(1) = 9.03$, $p = .003$, although it remained high ($Q(92) = 585.62$, $p < .001$, level 3: $\sigma^2 = 0.02$, 59.84%). For the funnel and forest plot, see Supplementary Material, Figs. S1 and S2.

3.2.2. Divergent validity

In total, we extracted 65 correlations reflecting divergent validity from 33 individual studies (see Table 1 for a detailed overview). Across disorders, medium correlations were most frequent while small and large correlations were equally frequent. This pattern was also found when focusing on disorders other than depression, for example SAD, GAD or Psychosis, where only few indices resembling divergent validity could be extracted. Finally, for depression small and medium correlations were most frequent with large correlations being slightly rarer.

3.2.3. Reliability

We extracted 53 indices of internal consistency from 31 studies (see Table 1 for a detailed overview). Results of our meta-analysis showed that the overall reliability coefficient in the intercept only model was good, $\alpha_{overall}$ [95%-CI] = 0.79 [0.74, 0.84], $SE = 0.03$ (see Table 1 for a detailed overview). However, the overall heterogeneity was high, $Q(52) = 1405.16$, $p < .001$, and the variance within studies was significant (level 2: $\sigma^2 = 0.003$, 14.31%), $\chi^2(1) = 101.90$, $p < .001$, as well as the variance between studies (level 3: $\sigma^2 = 0.019$, 84.54%), $\chi^2(1) = 28.46$, $p < .001$. Finally, 1.15% of the variance were accounted for by random sampling (level 1). Accordingly, we also conducted post-hoc exploratory moderation analyses for the internal consistency data. We found the application format to be a significant moderator, such that using a computerized version of the SST was associated with slightly higher reliability indices compared to studies using other application formats, $F(1, 51) = 5.35$, $p = .025$, with an estimated difference in $\alpha_{overall} = 0.12$ [0.02, 0.22], $SE = 0.05$. Further, comparing the disorders in which the SST was applied (depression vs. other), the reliability was slightly higher for depression, $F(1, 51) = 4.45$, $p = .04$, with a difference in the overall internal consistency of $\alpha_{overall} = 0.06$ [0.00, 0.12], $SE = 0.03$. However, language (English vs. other) was not a significant moderator, $F(1, 51) = 0.48$, $p = .493$, as well as the application of a cognitive load, $F(1, 51) = 0.62$, $p = .433$, the sample's diagnostic status (partly diagnosed participants versus non-diagnosed participants), $F(1, 51) = 0.14$, $p = .708$, or the number of trials, $F(1, 51) = 1.44$, $p = .235$. When comparing the intercept only model to the model containing the application format and the type of disorder as moderators of the overall reliability, heterogeneity was significantly reduced yet remained high, $\chi^2(1) = 9.14$, $p = .01$, ($Q(50) = 894.16$, $p < .001$, level 3: $\sigma^2 = 0.02$, 88.87%) (see supplementary material for the funnel and forest plot, Figs. S3 and S4).

Further, we extracted four reliability indices concerning test-retest reliability from three individual studies. Here, all indices fell within an acceptable range, ranging from $r = 0.76$ to ICC $= 0.93$ (see Table 1).

3.2.4. Risk of bias

We found the overall quality of the included studies to be high in relation to the data relevant for our review and there were only a few studies that presented unclear inclusion criteria, used a non-valid disorder assessment or did not report the percentage of missing data. However, across many studies researchers administering the SST may not have been blind to participants' symptom levels and/or group allocations (e.g., they may have been aware that the participant came from a clinical sample, or a paper-based symptom questionnaire may have been administered and collected by the researcher prior to administering the SST). Further, not many studies conducted formal power analyses to justify their sample sizes. Importantly, only a few studies reported on technical aspects of the SST such as the retention

---

[3] If the language in which the SST was applied was not explicitly mentioned, this information was derived from either the in-/exclusion criteria concerning language fluency or the language in which other tasks and questionnaires were applied.

[4] The high amount of between-study heterogeneity was likely not a result of us investigating convergent validity across disorders, as indicated by nearly identical heterogeneity when we restricted our analyses to depression only. See Supplementary Results for details.

[5] One study only reported a joint index for the SST with and without load and was thus excluded from this analysis.

[6] Except for two studies, the clinical studies, i.e., studies that included diagnosed participants who were compared to a non-diagnosed, non-clinical group, only reported correlations across the entire sample. As such, it was not possible to obtain separate correlation coefficients for diagnosed vs. non-diagnosed participants from these clinical studies. Hence, as a proxy for this moderator variable, we thus created a variable that differentiates between the samples from the clinical studies including diagnosed participants versus samples from the non-clinical studies fully consisting of non-diagnosed participants.

Table 1
Study characteristics, reliability & validity data.

Study	Type of sample (N)	Study design, SST type of application (sentences per block), scoring, and language	Reliability	Convergent validity	Divergent validity
				Low/Medium/High: k indices (% indices, n participants)	Low/Medium/High: k indices (% indices, n participants)
Bipolar Disorder I/II	N = 76		–	–	–
Miklowitz, Alatiq, Geddes, Goodwin, & Williams (2010)	Remitted Bipolar Disorder and MDD patients, and HC (76)	Cross-sectional, 4 SST blocks (29) Paper & Pencil, Negativity and Hyperpositivity score, English			
Borderline Personality Disorder (BPD)	N = 181			Low: 0 (0%) Medium: 2 (100%, *n* = 181) High: 0 (0%)	–
Geiger, Peters, & Baer (2014)	Students with varying levels of BPD features (181)	Cross-sectional, 2 SST blocks (25) Paper & Pencil, Positivity score, English	–	*Cognitive load:* PBQ-BPD: $r = -0.47$** *No cognitive load:* PBQ-BPD: $r = -0.45$**	–
Eating disorders	N = 103			Low: 0 (0%) Medium: 0 (0%) High: 5 (100%, *n* = 113)	Low: 1 (50%, *n* = 33) Medium: 0 (0%) High: 1 (50%, *n* = 37)
Bradatsch et al. (2020)	Women with elevated body dis-satisfaction (33)	Pre-Post, 1 SST block (20) at pre- and post, Computerized, Negativity score, German	–	*Pre-cue-exposure:* BISS pre-cue-exposure: $r = -0.79$***	–
Brockmeyer et al. (2018)	Female individuals with Anorexia (37) and HC (33) (Total 70)	Cross-sectional, 1 SST block (20) Computerized Negativity score, German	–	*Anorexia group:* BShQ: $r = 0.66$*** EDE-Q: $r = 0.57$*** *Healthy controls:* BShQ: $r = 0.54$*** EDE-Q: $r = 0.57$***	*Anorexia group:* PHQ-9: $r = 0.61$*** *Healthy controls:* PHQ-9: $r = 0.27$
Psychosis	N = 259		–	Low: 2 (15.38%, *n* = 70) Medium: 6 (46.15%, *n* = 198) High: 5 (38.46%, *n* = 201)	Low: 1 (11.11%, *n* = 70) Medium: 5 (55.56%, *n* = 198) High: 3 (33.33%, *n* = 140)
Savulich et al. (2020) Experiment 1	Healthy participants with differing levels of paranoid thinking (70)	Cross-sectional, 2 SST blocks (20), Paper & Pencil, Negativity score, English	Cronbach's α = 0.64	GPTS: $r = 0.47$*** PS: $r = 0.62$*** PDI: $r = 0.28$* CAPS: $r = 0.20$	BDI-II: $r = 0.45$*** STAI-T: $r = 0.58$***
Savulich, Freeman, Shergill, & Yiend (2015)	Students with varying levels of paranoid thinking (70)	Cross-sectional, 2 SST blocks (20) Paper & Pencil Negativity score, English	Split-Half = 0.83	*Paranoia related trials:* GPTS: $r = 0.47$*** PS: $r = 0.64$*** PDI: $r = 0.31$* CAPS: $r = 0.36$**	*Paranoia related trials:* BDI-II: $r = 0.47$*** DASS: $r = 0.49$*** STAI-T: $r = 0.67$*** PSWQ: $r = 0.52$*** CWI: $r = 0.12$ LSAS: $r = 0.44$***
Savulich, Shergill, & Yiend (2017)	Paranoid and non-paranoid schizophrenia patients, and HC (61)	Cross-sectional, 2 SST blocks (20), Paper & Pencil, Negativity score, English	Split-Half = 0.82	*Paranoia related trials:* GPTS: $r = 0.59$*** PS: $r = 0.55$*** PDI: $r = 0.74$***	–
Yiend et al. (2019)[a]	HC, at risk for psychosis and first episode psychosis patients (58)	Cross-sectional, 1 SST block (20), NA, Positivity score (Adaptive bias in the study), English	–	*Paranoia related trials: Adaptive bias:* GPTS: $r = -0.39$** PS: $r = -0.40$**	*Paranoia related trials: Adaptive bias:* DASS: $r = -0.39$**
Posttraumatic Stress Disorder (PTSD)	N = 270		–	Low: 0 (0%) Medium: 0 (0%) High: 1 (100%)	Low: 1 (25%, *n* = 214) Medium: 3 (75%, *n* = 214) High: 0 (0%)
Viviani, Mahler, et al. (2018) Behavioral study	Unselected students (36)	Cross-sectional, 1 SST block (48), Computerized Positivity score German	–	No zero order correlations presented	No zero order correlations presented
Viviani, Mahler, et al. (2018) Neuroimaging study	Unselected students (20)	Cross-sectional, 1 SST block (48), Computerized Positivity score German	–	No zero order correlations presented	No zero order correlations presented
*Würtz, Blackwell, Margraf, & Woud (2021)**	Participants with a distressing life event (214)		T1: Split-Half = 0.88	*All trials, T1:* PCL-5: $r = 0.51$***	*All trials, T1:* QIDS-SR: $r = 0.48$***

(continued on next page)

Table 1 (*continued*)

Study	Type of sample (N)	Study design, SST type of application (sentences per block), scoring, and language	Reliability	Convergent validity	Divergent validity
				Low/Medium/High: k indices (% indices, n participants)	Low/Medium/High: k indices (% indices, n participants)
		Longitudinal, 2 SST blocks (20), Computerized, Negativity score, German	T2: Split-Half = 0.90 Test-Retest (14 days): ICC = 0.81		EDEQ: $r = 0.30$*** FSS: $r = 0.24$*** AQ: $r = 0.32$***
Social Anxiety	*N = 478*	–	–	Low: 0 (0%) Medium: 1 (25%, $n = 60$) High: 3 (75%, $n = 119$)	Low: 0 (0%) Medium: 1 (100%, $n = 60$) High: 0 (0%)
Bowler et al. (2012)	Socially anxious students (63)	Pre-Post, 2 SST blocks (20) pre and post, Paper & Pencil, Negativity score, English	–	–	–
Burnett Heyes et al. *(2017) Pilot study* cited in Burnett Heyes* et al. *(2017)*	Unselected young adults (60)	Cross-sectional, 1 SST block (20), Paper & Pencil, Negativity score, English	–	LSA: $r = 0.35$**	STAI-T: $r = 0.41$***
Burnett Heyes et al. (2017)	Unselected adolescents (57)	Pre-Post, 1 SST block (20) post and at follow-up, Paper & Pencil, Negativity score, English	–	SST assessed after CBM paradigm	SST assessed after CBM paradigm
de Voogd et al. (2017)	Adolescents with elevated anxiety/depression (119)	Pre-Post, 3 SST blocks (10) pre and post, Computerized, Negativity score, Dutch	Set 1: Split-Half = 0.87 Set 2: Split-Half = 0.71	*Depression & social anxiety related trials:* SCARED: $r = 0.62$*** CDI: $r = 0.75$***	–
Krans, Bosmans, Salemink, & De Raedt (2019)	Unselected students (120)	Pre-Post, 1 SST block (20) at post, Computerized, Positivity score, Dutch	Social: Cronbach's α = 0.88 Future related: Cronbach's α = 0.82	SST assessed after CBM paradigm	SST assessed after CBM paradigm
Standage (2009) pilot study* cited in Standage, Ashwin, & Fox (2010)	N/A	Cross-sectional, NA (NA), N/A, Positivity score, English	–	FNE: $r = -0.64$***	–
Standage, Ashwin, & Fox (2010) Experiment 1	Unselected students (30)	Pre-Post, 1 SST block (20) pre and post, Paper & Pencil, Positivity score, English	–	–	–
Experiment 2	Unselected students (29)	Pre-Post, 1 SST block (20) pre and post, Paper & Pencil, Positivity score, English	–	–	–
Generalized Anxiety Disorder (GAD)	*N = 1132*	–	–	Low: 4 (36.36%, $n = 343$) Medium: 2 (18.18%, $n = 77$) High: 6 (45.45%, $n = 546$)	Low: 0 (0%) Medium: 3 (75%, $n = 420$) High: 1 (25%, $n = 88$)
Deschênes, Dugas, Anderson, & Gouin (2015)	Unselected students (77)	Pre-Post, 1 SST block (20) pre and post, NA, Negativity score, English	–	PSWQ: $r = 0.30$** STICSA-T: $r = 0.34$**	CES-D: $r = 0.42$***
Donegan (2016) Dissertation*	Individuals with GAD (29)	Pre-Post with 1-week follow-up, 1 SST block (20) per session, Computerized, Negativity score French	–	–	–
Hirsch et al. (2018)[b]	MDD and GAD patients (130)	Pre-Post, 1 SST block (20) pre and post, NA, Positivity score, English	–	–	–
Hirsch et al. (2021)	Community volunteers with GAD (183)	Pre-Post, 1 SST block (20) pre and post, Computerized,	Set 1, Pre: Cronbach's α = 0.77	PSWQ: $r = -0.08$ GAD-7: $r = -0.24$**	PHQ-9: $r = -0.41$**

(*continued on next page*)

Table 1 (*continued*)

Study	Type of sample (*N*)	Study design, SST type of application (sentences per block), scoring, and language	Reliability	Convergent validity	Divergent validity
				Low/Medium/High: k indices (% indices, n participants)	Low/Medium/High: k indices (% indices, n participants)
		Positivity score, English	Set 2, Pre: Cronbach's α = 0.79 Set 1, Post: Cronbach's α = 0.82 Set 2, Post: Cronbach's α = 0.88		
Hirsch et al. (2020)	Community sample with elevated rumination/worry (160)	Pre-Post, 1 SST block (20) pre and post, Computerized, Positivity score, English	–	*Worry related trials:* PSWQ: *r* = −0.20* GAD-7: *r* = −0.17*	*Worry related trials:* PHQ-9: *r* = −0.30**
Hirsch, Meeten, et al. (2020)	Community sample of pregnant and non-pregnant women (237)	Cross-sectional, 1 SST block (20), Computerized, Positivity score, English	Cronbach's α = 0.86	*Full sample: Depression and worry related trials:* PSWQ: *r* = −0.67*** PHQ-9: *r* = −0.62*** GAD-7: *r* = −0.63***	–
*Krahé, Meeten, & Hirsch (2020) *	Community sample (88)	Cross-sectional with re-test after 2 weeks, 1 SST block (20) per session, Computerized, Positivity score, English	Test-retest (14 days): ICC = 0.89 List 1, time 1: Cronbach's α = 0.77 List 2, time 1: Cronbach's α = 0.82 List 1, time 2: Cronbach's α = 0.89 List 2, time 2: Cronbach's α = 0.87	*Worry related trials:* PSWQ: *r* = −0.64** GAD-7: *r* = −0.64**	*Worry related trials:* PHQ-9: *r* = −0.56**
Krahé, Whyte, Bridge, Loizou, & Hirsch (2019)[b]	Individuals with GAD or MDD, and HC (221)	Cross-sectional, 1 SST blocks (20), Paper & Pencil, Positivity score, English	–	PSWQ: *r* = −0.72**	–
*McNally (2014) Dissertation**	GAD patients (7)	Pre-Post-Follow-up, 1 SST block at baseline, 1 pre-CBM, 1 post-CBM, 1 at follow-up (1 week), Paper & Pencil, Negativity score, English	–	–	–
Panic Disorder	N = 54	–	–	Low: 0 (0%) Medium: 1 (100%) High: 0 (0%)	Low: 0 (0%) Medium: 1 (33.33%, *n* = 54) High: 2 (67.67%, *n* = 54)
Zahler et al. (2020)	Students with differing levels of anxiety sensitivity (54)	Cross-Sectional, 1 SST block (24), Computerized, Negativity score, German	Split-Half = 0.73	BSeQ: *r* = 0.38**	DASS: *r* = 0.57** STAI-S: *r* = 0.47** STAI-T: *r* = 0.54**
Female Sexual Dysfunction	N = 263	–	–	Low: 1 (50%) Medium: 0 (0%) High: 1 (50%)	Low: 1 (100%, *n* = 263) Medium: 0 (0%) High: 0 (0%)
Zahler et al. (2021)	Female community sample (263)	Cross-Sectional, 1 SST block (20), Computerized, Negativity score, German	Split-Half = 0.86	*All participants, all trials:* SIDI-F: *r* = −0.56*** FSDS: *r* = 0.29***	*All participants, all trials:* PHQ-9: *r* = 0.18**
Depression	N = 7534	–	–	Low: 16 (28.57%, *n* = 1101) Medium: 22 (39.29%, *n* = 1881)	Low: 15 (36.59%, *n* = 1341) Medium: 14 (34.15%, *n* = 1825)

Table 1 (*continued*)

Study	Type of sample (N)	Study design, SST type of application (sentences per block), scoring, and language	Reliability	Convergent validity	Divergent validity
				Low/Medium/High: k indices (% indices, n participants)	Low/Medium/High: k indices (% indices, n participants)
				High: 18 (32.14%, n = 2428)	High: 12 (29.27%, n = 487)
*Beevers (2002) Dissertation**	Remitted and never depressed students (116)	Pre-Post, with 6–8 weeks follow-up and 50 days follow-up, 1 SST block (15) pre and post, Computerized/Pen & Paper, Negativity score, English	*No cognitive load:* Test-Retest (50 days): r = 0.76	*No cognitive load:* BDI-II: r = 0.51***	–
Beevers & Meyer (2008)	Non-depressed students (65)	Post manipulation, 1 SST block (60), Computerized/Pen & Paper, Negativity score, English	–	SST assessed after an experimental manipulation	–
Bibi, Margraf, & Blackwell (2020)	Depressed students (55) Intention to treat sample	Pre-Post-Follow-Up, 1 SST block (20) pre and post, Paper & Pencil, Negativity score, English	Pre, Set A: Split-Half = −0.21 Pre, Set B: Split-Half = −0.35 Post, Set A: Split-Half = 0.30 Post, Set B: Split-Half = −0.13	QIDS-SR: r = −0.03 DASS-Depression: r = 0.05	DASS-Anxiety: r = 0.13
Blackwell & Holmes (2010)	Depressed individuals (7)	Baseline-Pre-Post, 1 SST block (20), baseline, pre and post, Paper & Pencil, Negativity score, English	Baseline: Split-Half = 0.72 Pre: Split-Half = 0.64 Post: Split-Half = 0.87	BDI-II: r = 0.06	–
Blackwell et al. (2015); Ji, Holmes, & Blackwell (2017); Blackwell & Holmes (2017)	Depressed individuals (150)	Pre-Post, 1 SST block (20) pre and post, Paper & Pencil, Negativity score, English	Pre: Split-Half = 0.83 Post: Split-Half = 0.82	BDI-II: r = 0.51***	–
Blanco, Boemo, & Sanchez-Lopez (2021)	Unselected adults (80)	Cross-sectional, 1 SST block (15), Computerized, Negativity score, Spanish	Split-Half = 0.86	*No cognitive load* CES-D: r = 0.62***	GAD-7: r = 0.49***
Brockmeyer et al. (2012)	Remitted and never depressed (40)	Pre-Post, 1 SST block (20) pre and post, Paper & Pencil, Negativity score, German	–	*No cognitive load* BDI-II: r = 0.27 HDRS: r = 0.13	–
Cowden Hindash & Rottenberg (2017a)	Dysphoric and non-dysphoric students (160)	Cross-sectional, 1 SST block (20), Paper & Pencil, Negativity score, English	Split-Half = 0.90	BDI-II: r = 0.47***	STAI-S: r = 0.27*** STAI-T: r = 0.39***
*Cowden Hindash & Rottenberg (2017b)**	Dysphoric and non-dysphoric students (162)	Pre-Post, 1 SST block (20) at post, Paper & Pencil, Negativity score, English	Split-Half = 0.92	SST assessed after a CBM paradigm	SST assessed after a CBM paradigm
*Dickson (2015) Dissertation**	Remitted depressed & never depressed (38)	Cross-sectional, 2 SST blocks (20), Paper & Pencil, Negativity score, English	–	*Joint index for load- and no load:* BDI-II: r = 0.37*	–
Everaert, Mogoaşe, David, & Koster (2015) Experiment 2a[c]	Non-depressed students (38)	Pre-Post, 1 SST block (24) pre and post, N/A, Negativity score, Dutch	–	BDI-II: r = 0.41*	STAI-T: r = 0.50**
Experiment 2b	Non-depressed students (39)	Pre-Post, 1 SST block (24) pre and post, N/A, Negativity score, Dutch	–	–	–
Everaert, Tierens, Uzieblo, & Koster (2013)	Students with varying levels of depressive symptoms (64)	Cross-sectional, 1 SST block (20), Computerized, Negativity score, Dutch	–	BDI-II: r = 0.77***	–
Everaert, Duyck, & Koster (2014)	Students with varying levels of depressive symptoms (71)	Cross-sectional, 10 SST blocks (10), Computerized, Negativity score, Dutch	Split-Half = 0.86	BDI-II: r = 0.61***	–
Everaert, Grahek, Duyck, et al. (2016) Everaert, Grahek, & Koster (2016)	Students with varying levels of depressive symptoms (112)	Cross-sectional, 4 SST block (10), Computerized/Verbal, Negativity score Dutch	Split-Half = 0.82	BDI-II: r = 0.49***	DASS Anxiety: r = 0.29**
Gill, Miller, Haner, & Rude (2017) Study 2	Remitted depressed (53)	Cross-sectional, 2 SST blocks (7), Paper & Pencil, Proportion of Big-Picture-Appraisal related sentences, English	–	CES-D: r = −0.43***	–
Hedlund & Rude (1995)			–		–

Table 1 (*continued*)

Study	Type of sample (*N*)	Study design, SST type of application (sentences per block), scoring, and language	Reliability	Convergent validity	Divergent validity
				Low/Medium/High: k indices (% indices, n participants)	Low/Medium/High: k indices (% indices, n participants)
	Patients, remitted depressed and HC (53)	Cross-sectional, 2 SST blocks (30), Computerized / Paper & Pencil, Negativity score, English		SST assessed after a self-focus manipulation	
Holmes, Lang, & Shah (2009), Experiment 1	Community sample (38)	Pre-Post, 1 SST block (20) post, Paper & Pencil, Negativity score, English	–	–	–
Hirsch et al. (2020)	Community sample with elevated rumination/worry (160)	Pre-Post, 1 SST block (20) pre and post, Computerized, Positivity score, English	–	*Depression related trials* PHQ-9: $r = -0.31$**	*Depression related trials* PSWQ: $r = -0.12$ GAD-7: $r = -0.21$**
*Krahé, Meeten, & Hirsch (2020) **	Community sample (88)	Cross-sectional with re-test after 2 weeks, 1 SST block (20) per session, Computerized, Positivity score, English	Test-retest (14 days): ICC = 0.93 List 1, time 1: Cronbach's α = 0.92 List 2, time 1: Cronbach's α = 0.92 List 1, time 2: Cronbach's α = 0.96 List 2, time 2: Cronbach's α = 0.97	*Depression related trials:* PHQ-9: $r = -0.62$**	*Depression related trials:* PSWQ: $r = -0.64$** GAD-7: $r = -0.61$**
Krejtz, Holas, Rusanowska, & Nezlek (2018)	Depressed individuals (51 for SST data)	Pre-Post, 1 SST block (20) pre and post, Computerized, Positivity score, Polish	Split-half = 0.86	CES-D: $r = -0.33$** HADS-Depression: $r = -0.42$***	STAI-T: $r = -0.53$*** HADS-Anxiety: $r = -0.20$
Lang, Blackwell, Harmer, Davison, & Holmes (2012)	Depressed individuals (26)	Pre-Post, 1 SST block (20) pre and post, Computerized, Negativity score, English	–	BDI-II: $r = 0.53$** HDRS: $r = 0.44$*	STAI-T: $r = 0.43$*
Lee, Mathews, Shergill, & Yiend (2016)	Depressed individuals (71)	Cross-sectional, 1 SST block (20), Paper & Pencil, Negativity score, English	–	BDI-II: $r = 0.33$**	–
LeMoult et al. (2018)	Depressed adolescents (46)	Pre-Post, 1 SST block (20) pre and post, N/A, Positivity score, English	–	–	–
Lin et al. (2020) Study 1	Unselected students (145)	Cross-sectional, 4 SST blocks (10), Computerized, Negativity score, Chinese	Split-Half = 0.63	PHQ-9: $r = 0.30$***	GAD-2: $r = 0.22$**
*Lo (2009) Dissertation**	Remitted and never depressed (40)	Cross-sectional, 6 SST blocks (20), Computerized, Negativity score German	–	–	–
Marchetti, Koster, & De Raedt (2012)	Unselected students (79)	Pre-Post, 1 SST block (20) pre and post, Paper & Pencil, Negativity score Dutch	–	*Mind wandering induction between BDI-II and SST*	–
*Martin-Romero, Sanchez-Lopez, & Vazquez (in prep), Study 1**	Remitted and never depressed (109)	Cross-sectional, 2 SST blocks (20), Paper & Pencil, Negativity score, Spanish	–	*Cognitive load:* BDI-II: $r = 0.41$** DID: $r = 0.30$*** *No cognitive load:* BDI-II: $r = 0.39$** DID: $r = 0.28$***	*Cognitive load:* BAI: $r = 0.27$*** *No cognitive load:* BAI: $r = 0.28$***
*Martin-Romero, Sanchez-Lopez, & Vazquez (in prep), Study 2**	Depressed, remitted, never depressed (409)	Cross-sectional, 1 SST block (20), Paper & Pencil, Negativity score, Spanish	–	BDI-II: $r = 0.38$***	BAI: $r = 0.28$***
*Martin-Romero, Sanchez-Lopez, & Vazquez (in prep), Study 3**	Currently and never depressed (80)	Cross-sectional, 1 SST block (20), Paper & Pencil, Negativity score, Spanish	–	BDI-II: $r = 0.65$***	BAI: $r = 0.43$***
Miller, Rude, & Haner (2015)	Unselected students (137)	Pre-Post, 2 SST blocks (8) post, Computerized, Training congruent vs. incongruent sentences, English	–	SST assessed after an experimental manipulation	–
Moon & Lord (2006), Study 1	Unselected students (91)	Cross-sectional, 1 SST block (40), Paper & Pencil, Negativity score, English	–	SST assessed after an experimental manipulation	–

Table 1 (*continued*)

Study	Type of sample (*N*)	Study design, SST type of application (sentences per block), scoring, and language	Reliability	Convergent validity	Divergent validity
				Low/Medium/High: k indices (% indices, n participants)	Low/Medium/High: k indices (% indices, n participants)
Moon & Lord (2006), Study 3	Unselected students (100)	Cross-sectional, 1 SST block (40), Paper & Pencil, Negativity score, English	–	SST assessed after an experimental manipulation	–
Murphy et al. (2015)	Healthy older adults (77)	Pre-Post-Follow-up, 1 SST block (20) pre post, and follow-up, Paper & Pencil, Negativity score, English	–	BDI-II: $r = 0.34$**	STAI-T: $r = 0.43$**
Novović, Mihić, Biro, & Tovilović (2014)	Unselected students (1071)	Cross-sectional, 2 SST blocks (28), Paper & Pencil, Negativity score, Serbian	*Cognitive load* Split-Half = 0.62 *No cognitive load:* Split-Half = 0.64	*Joint index for load- and no load:* DASS-Depression: $r = 0.54$***	*Joint index for load- and no load:* DASS-Anxiety: $r = 0.35$***
O'Connor, Everaert, & Fitzgerald (2021)	Unselected students (82)	Cross-sectional, 1 SST block (60), Computerized, Negativity score, English	Split-Half = 0.95	BDI-II: $r = 0.80$***	STAI-T: $r = 0.64$***
Penton-Voak et al. (2020) Study 2	Depressed individuals (190)	Post-Follow-Up-Follow-Up, 1 SST block (20) per assessment, N/A, Negativity score, English	–	SST assessed after an experimental manipulation	–
Phillips, Hine, & Bhullar (2012)	Unselected students (306)	Cross-sectional, 1 SST block (20), Computerized, Negativity score, English	–	SST assessed after sad mood induction	–
Phillips and Hine (2013)	Unselected students (355)	Cross-Sectional, 1 SST block (20), Computerized, Negativity score, English	–	SST assessed after sad mood induction	–
Rohrbacher (2016) *dissertation**	Students with varying levels of depressive symptoms (70)	Cross-Sectional, 1 SST block (20), Paper & Pencil, Negativity score German	Cronbach's alpha = 0.57	BDI-II: $r = 0.47$**	–
Romero, Sanchez, & Vazquez (2014)	Remitted and never depressed students (70)	Cross-sectional, 1 SST block (20), Paper & Pencil, Negativity score, Spanish	–	BDI-II: $r = 0.25$* DID: $r = 0.19$	BAI: $r = 0.11$
Rude, Covich, Jarrold, Hedlund, & Zentner (2001)	Remitted patients and never depressed (41)	Pre-Post, 2 SST blocks (20) at post, Computerized, Negativity score, English	–	SST assessed after experimental manipulation	–
Rude, Durham-Fowler, Baum, Rooney, & Maestas (2010); Maestas et al. (2008)	Remitted or never depressed women (73)	Longitudinal (6–18 months), 2 SST blocks (25) at first session, Paper & Pencil, Positivity score, English	*Cognitive load:* Split-Half = 0.53	*Cognitive load:* BDI-II: $r = -0.43$* *No cognitive load:* BDI-II: $r = -0.29$**	–
*Rude, Durham-Fowler, Baum, Rooney, & Maestas, (2010); cited in Maestas et al. (2008)**	Male partners of women from the same study (73)	Longitudinal (6–18 months), 2 SST blocks (25) at first session, Paper & Pencil, Positivity score, English	*Cognitive load:* Split-Half = 0.61	*Cognitive load:* BDI-II: $r = -0.23$	–
Rude, Mazzetti, Pal, & Stauble (2011)	Students with a recent rejection experience (196)	Pre-Post, 1 SST block (25) at post, Paper & Pencil, Negativity score, English		SST assessed after experimental manipulation	–
Rude, Wenzlaff, Gibbs, Vane, & Whitney (2002); Rude, Valdez, Odom, & Ebrahimi (2003)	Unselected Students (339), Subsample of the first sample tested again (84)	Longitudinal (4–6 weeks, and 18–28 months for the subsample), 2 SST blocks (20) at first session, Paper & Pencil, Negativity score, English	–	*Cognitive load:* BDI-II: $r = 0.51$** IDD: $r = 0.25$** *No cognitive load:* BDI-II: $r = 0.55$** IDD: $r = 0.27$**	–
Sanchez, Duque, Romero, & Vazquez (2017)	Currently and never depressed (58)	Cross-sectional, 1 SST block (20), Computerized, Negativity score, Spanish	–	BDI-II: $r = 0.62$***	BAI: $r = 0.44$***
Sanchez, Everaert, De Putter, Mueller, & Koster (2015)	Students with varying levels of depressive symptoms (52)	Cross-sectional, 5 SST blocks (8), Computerized, Positivity score, Dutch	Split-Half = 0.88	*No valence instruction: No cognitive load:* BDI-II: $r = -0.75$**	STAI-T: $r = 0.44$***
Sanchez, Everaert, & Koster (2016)	Students with varying levels of depressive symptoms (36)	Cross-sectional, 1 SST block (21), Computerized, Negativity score for correlations (Positivity in the paper), Dutch	Split-Half = 0.89	*Before manipulation: No cognitive load:* BDI-II: $r = -0.60$*** DASS-Depression: $r = -0.51$***	DASS-Anxiety: $r = -0.42$**
Sanchez-Lopez, De Raedt, van Put, & Koster (2019)	Unselected students (41)	Cross-sectional, 1 SST block (12), Computerized, NA, NA	–	–	–
Sanchez-Lopez, Everaert, Van Put, De Raedt, & Koster (2019)	Unselected students (66)	Cross-sectional, 2 SST blocks (18), Computerized, NA, NA	–	–	–
Savulich, Freeman, Shergill, & Yiend (2015)	Students with varying levels of paranoid thinking (70)	Cross-sectional, 2 SST blocks (20) Paper & Pencil Negativity score, English	–	*Generally emotional trials:* BDI-II: $r = 0.34$**	*Generally emotional trials:* STAI-T: $r = 0.53$*** GPTS: $r = 0.27$*

Table 1 (*continued*)

Study	Type of sample (N)	Study design, SST type of application (sentences per block), scoring, and language	Reliability	Convergent validity	Divergent validity
				Low/Medium/High: k indices (% indices, n participants)	Low/Medium/High: k indices (% indices, n participants)
Savulich, Shergill, & Yiend (2017)	Paranoid and non-paranoid schizophrenia patients, and HC (61)	Cross-sectional, 2 SST blocks (20), Paper & Pencil, Negativity score, English	–	–	PDI: $r = 0.16$ CAPS: $r = 0.21$ PS: $r = 0.40**$ PSWQ: $r = 0.37**$ LSAS: $r = 0.33*$ CWI: $r = 0.25*$ *Generally emotional trials:* GPTS: $r = 0.61***$ PDI: $r = 0.60***$ PS: $r = 0.55***$
*Seeds (2012) Dissertation**	Unselected female students (180)	Longitudinal (5 weeks), 2 SST blocks (20) at first time point, Computerized, Negativity score, English	–	*Cognitive load:* BDI-II: $r = 0.42***$ IDD-L: $r = 0.22***$ *No cognitive load:* BDI-II: $r = 0.43***$ IDD-L: $r = 0.20**$	–
Sfärlea et al. (2019)$_d$	Parent (remitted & never depressed) -Children Dyads (156 (78 dyads))	Cross-sectional, Parents: 7 SST blocks (10) Children: 5 SST blocks (10), Computerized, Negativity score German	Children: Split-Half = 0.53 Adults: Split-Half = 0.78	SST assessed after a mood manipulation	–
Sfärlea et al. (2020)$_d$; (Sfärlea et al., 2021)	Children and adolescents, either depressed, or at high or low familial risk for depression (121)	1 SST block (50), Computerized, Negativity score, German	Split-Half = 0.94	SST assessed after a mood manipulation	SST assessed after a mood manipulation
Torkan et al. (2014)	Depressed, patients (39)	Pre, Post, Follow-Up, 1 SST block (20) per session, Paper & Pencil, Negativity score, Persian	–	BDI-II: $r = 0.31$	STAI-T: $r = 0.59***$
Van der Does (2005)	Remitted patients and never depressed individuals (46)	Cross-sectional, 3 SST blocks (20), Computerized, Negativity score, Dutch	–	SST assessed after a mood manipulation	–
Van der Does, Manthey, & Hermans (2012)	Remitted and never depressed individuals (64)	Pre-Post, 2 SST blocks (20) at post, Paper & Pencil, Negativity score Dutch	–	After a mood manipulation/thought suppression task	–
Viviani, Dommes, Bosch, Stingl, & Beschoner (2018) Behavioral study	Unselected students (43)	Cross-sectional, 1 SST block (80), Computerized, Positivity score, German	–	No zero order correlations presented	No zero order correlations presented
Viviani et al. (2010)	Unselected students (60)	Cross-sectional, 6 SST blocks (20–22) Paper & Pencil / Computerized Negativity score, German	–	No zero order correlations presented	No zero order correlations presented
Watkins & Moulds (2007)	Currently or remitted depressed patients and HC (60)	Cross-sectional, 2 SST blocks (20), Paper & Pencil, Negativity score, English	–	–	–
Wells, Vanderlind, Selby, & Beevers (2014)	Non-depressed students (155)	Longitudinal (6 months), 2 SST blocks (20) at first session, Paper & Pencil, Negativity score, English	–	*Cognitive load:* CES-D: $r = 0.14$ *No cognitive load:* CES-D: $r = 0.07$	–
Wenzlaff and Bates (1998)	Non-depressed, at-risk and depressed students (90)	Cross-sectional, 3 SST blocks (20), Paper & Pencil, Negativity score, English	–	–	–
Westermann et al. (2021)	Patients (57)	Pre-Post, 1 SST block (20) Pre and Post, Paper & Pencil, Negativity score German	Pre: Split-Half = 0.90 Post: Split-Half = 0.87	QIDS-SR: $r = 0.58$	GAD-7: $r = 0.40$
Williams, Blackwell, Mackenzie, Holmes, & Andrews (2013)	Depressed individuals (69)	Pre-Post, NA SST block (NA) Pre and Post, Paper & Pencil, Computerized, Negativity score, English	–	–	–
Yiend et al. (2014)	Depressed individuals (36)	Pre-Post, 1 SST block (20) pre and post,	–	Stressor task for a subset of participants before SST assessment	–

Table 1 (*continued*)

Study	Type of sample (*N*)	Study design, SST type of application (sentences per block), scoring, and language	Reliability	Convergent validity	Divergent validity
				Low/Medium/High: k indices (% indices, n participants)	Low/Medium/High: k indices (% indices, n participants)
Yiend et al. (2019)a	HC, at risk for psychosis and first episode psychosis patients (58)	Paper & Pencil, Negativity score, English Cross-sectional, 1 SST block (20), Paper & Pencil, Positivity score, English	–	*Generally emotional trials:* *Adaptive bias:* DASS-Depression: $r = -0.60$***	*Generally emotional trials:* *Adaptive bias:* DASS-Anxiety: $r = -0.54$*** GPTS: $r = -0.45$*** PS: $r = -0.51$***
Overall		–		Low: 23 (24.21%, n = 1777) Medium: 34 (35.79%, n = 2381) High: 38 (40.00%, n = 3751)	Low: 19 (29.23%, n = 1851) Medium: 27 (41.54%, n = 2643) High: 19 (29.23%, n = 648)

Note. Studies with the authors' names in italics and marked with an asterisk indicate non-peer reviewed studies. a = positivity and negativity scores were presented with differing values, only one index was extracted in this table. b, d = Samples were partly drawn from larger studies; thus, some participants are included in both studies. Studies with the same subscript are derived from the same larger study. c = Raw data was provided, yet due to the number of erroneous trials which we coded as missing values, we could not estimate split-half reliability. HC: Healthy Controls (for details on sample descriptions, see Supplementary Material). AQ: Acrophobia Questionnaire – Anxiety subscale. BAI: Beck Anxiety Inventory. BDI(-II): Beck Depression Inventory (-II). BISS: Body Image State Scale. BShQ: Body Shape Questionnaire. BSeQ: Body Sensations Questionnaire. CAPS: Cardiff Anomalous Perceptions Scale. CDI: Children's Depression Inventory. CES-D: Center for Epidemiological Studies – Depression Scale. DARS: Dimensional Anhedonia Rating Scale. DASS: Depression, Anxiety, Stress Scales. DID: Diagnostic Interview for Depression. EDE-Q: Eating Disorder Examination Questionnaire. FNE: Fear of Negative Evaluation Scale. FSS: Fear of Spiders Screening. GAD-2/7: Generalized Anxiety Disorder 2/7-item. GPTS: Green Paranoid Thoughts Scale. HADS: Hospital Anxiety and Depression Scale. HDRS: Hamilton Depression Rating Scale. IDD: Inventory to Diagnose Depression. LSA: Liebowitz Social Anxiety Scale. PAI-BOR: Personality Assessment Inventory – Borderline Features Scale. PCL-5 = PTSD Checklist for DSM-5. PDI: Peter's Delusions Inventory. PHQ-9: Patient Health Questionnaire. PS: Paranoia Scale. PSWQ: Penn's State Worry Questionnaire. QIDS-SR: Quick Inventory of Depressive Symptomatology. SCARED: Screen for Child Anxiety Related Disorders. SIDI-F: Sexual Interest and Desire Inventory - Female. STAI-S/T: State Trait Anxiety Inventory (State/Trait version).

Table 2
Overall ROB across studies.

Risk of bias item	% of studies fulfilling the criterion
1. Research question clear	100.00
2. Included population clear	98.91
3. Inclusion criteria clear	98.90
4. Power analysis conducted	18.18
5. Variance in symptoms of the targeted disorder	94.51
6. Valid disorder measure	100.00
7. Blinded SST application	23.86
8. Error rate on the SST reported	23.86
9. Load retention reported in studies applying a cognitive load	19.19
10. Excluded participants reported with reasons	95.60

rate of the cognitive load or the number of excluded trials due to (grammatical) errors. For a summary, see Table 2, for information on the ROB of individual studies, see Table S2 in the Supplementary Material.

4. Discussion

The aim of the current review was twofold: (i) providing a systematic overview of the SST's application, including e.g., types of samples and disorder domains, and (ii) evaluating the SST's psychometric properties, i.e., its validity and reliability, from a meta-analytic perspective. In the following, we will present a summary of our main results and their implications and provide some directions for future research.

4.1. Samples, disorder domains, and application characteristics

The SST has been applied mostly in undiagnosed samples, such as unselected student samples or samples including individuals with elevated symptom levels. About a third of the included studies applied the SST in a clinical study setting, testing samples that e.g., included

diagnosed or remitted patients, or recruited participants from healthcare centers. This distribution emphasizes the applicability of the SST across different disorder severities. This is further supported by our moderation analysis, indicating that the SST's psychometric properties did not differ between the different types of samples. While the SST was initially developed for research on IBs in depression, it is now applied more broadly, mostly in the context of GAD and SAD, but also in the context of Bipolar Disorder (I/II), Borderline Personality Disorder, Eating Disorders, Female Sexual Dysfunction, Panic Disorder, Psychosis, and Posttraumatic Stress Disorder. Further, the SST has been used in different age ranges, i.e., in older adults, children, and adolescents. According to our results there are two main application formats, fully computerized or pencil & paper application. While most of the studies applied the SST in English, the SST has also been used in Chinese, Dutch, Persian, French, German, Polish, Serbian, and Spanish.

4.2. Convergent validity

Results of our meta-analysis suggest that the SST is moderately to strongly correlated with symptoms of the corresponding target disorder, indicating good convergent validity. However, this conclusion needs to be treated with caution due to the high level of heterogeneity between the included studies. Via exploratory post-hoc analyses, we tried to better understand this heterogeneity and tested the moderating effects of a number of variables. Here, results showed that application format, but not the use of a cognitive load, was a significant moderator, with better validity indices for computerized SST versions. Further, the SST's convergent validity depended neither on the type of sample included nor on the targeted disorder. While these results do not explain the studies' heterogeneity, they do underline the broad applicability of the SST.

4.3. Divergent validity

For divergent validity indices, we found mostly medium size

correlations, with small correlations being relatively rare. When examining trait anxiety as a correlate, only medium to large correlations were found. While this potentially suggests that the SST is a rather unspecific measure of IBs, this needs to be interpreted cautiously. First, although small correlations were rare, the studies generating these indices included larger samples compared to those studies finding large correlations. Hence, large correlations may be overrepresented. Another important issue relates to the medium to large correlations between the SST and trait anxiety. Trait anxiety was assessed via the State-Trait Anxiety Inventory (Trait version; STAI-T; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). A recent meta-analysis, however, has indicated that the STAI-T could be also interpreted as a measure of negative affectivity, rather than a specific measure of trait anxiety (Knowles & Olatunji, 2020). Hence, our findings may thus represent an association between the broader concept of negative affectivity and IBs, instead of a lack of specificity of the SST.

Another reason for our mixed findings concerning divergent validity could be that the abundance of medium sized correlations might represent a methodological artifact of the general design of the studies included. That is, there were only a few formal validation studies including a heterogeneous questionnaire battery also assessing symptoms of disorders not directly linked to the targeted one. Hence, the reported correlations may represent comorbidity rather than a lack of divergent validity. This is further supported by the result that large correlations in some symptom domains were found in studies including patients but not in healthy samples. For example, in anorexia nervosa, the correlation with depression was large for participants with anorexia but small for healthy controls (Brockmeyer et al., 2018), and in relation to depression, correlations with paranoid thinking and delusions were medium or large when the sample consisted partly of patients or participants at risk (Savulich, Shergill, & Yiend, 2017; Yiend et al., 2019), but small when the sample consisted of students (Savulich et al., 2015). As such, the question of whether the SST is capable of assessing disorder specific IBs in the context of comorbidities remains a key challenge for future research.

4.4. Reliability

Results of our meta-analysis results showed that overall, the SST's internal consistency was good. However, similar to our conclusions regarding the convergent validity indices, this conclusion concerning the SST's reliability should be treated with caution due to high levels of heterogeneity. Via our exploratory post-hoc moderation analyses, two potential sources of the studies' heterogeneity could be identified statistically: First, the application format, with slightly higher reliability indices for computerized compared to other formats. Second, the reliability of the SST was slightly higher when applied in depression compared to other disorders. However, none of the other moderators were significant, with the lack of significance for the number of trials being one of the most surprising findings. Another important note when interpreting the overall internal consistency is the computational process of the reliability indices itself. Only a few studies reported reliability indices, especially for test-retest reliability, and we supplemented the published data with our own calculations based on data provided by the studies' authors. As a consequence, our reliability indices cannot be generalized and are only representative of a certain small set of studies. Hence, and in line with Parsons et al. (2019), we strongly recommend that if computationally possible, the split-half reliability should always be reported, and in case this index is not computable (e.g., due to a high ratio of missing/erroneous trials), this should itself be reported and explained.

Considering re-test reliability, we could only extract four indices (one for GAD, one in Posttraumatic Stress Disorder, two for depression). While these indices indicated sufficient reliability for a time period of two weeks and 50 days, the generalizability of these results is clearly limited. Therefore, more research is needed investigating the SST's test-

rest reliability, to ensure that experimental, clinical, and longitudinal studies can apply the SST as a reliable tool to investigate IBs, and to ensure findings are not distorted by random fluctuations.

4.5. The SST in context of other measures of IBs

Our findings can be integrated into the broader literature on the psychometric properties of IB measures (a summary of such measures can be found in Table S1 in the Supplementary Material). A useful reference point is provided by the systematic review by Gonsalves, Whittles, Weisberg, and Beard (2019), which investigated the psychometric properties of another well-established measure of IBs, the Word-Sentence-Association Paradigm (WSAP; Beard & Amir, 2009). Across various disorders, Gonsalves et al. (2019) found that small correlations for convergent validity were most common while large correlations were rare; this contrasts with the moderate to large correlations more common for the SST as found in the present review. However, findings for reliability of the WSAP and SST appear comparable, with Gonsalves et al. (2019) noting that acceptable reliability indices were most frequent for the WSAP, but also that this information was seldom reported. To the best of our knowledge, the present review and the review by Gonsalves et al. are the only published systematic overviews of the psychometric properties of measure of IBs across different types of disorders. Within a disorder-specific context, Everaert et al. (2017) investigated the relationship between IB measures, including the SST, and depression symptoms. They found an overall large association between measures of IB and symptoms, but only when using either self-report measures or the SST, and not when using measures relying fully on reaction times, for example semantic priming paradigms.

Although a meta-analysis such as that by Everaert et al. (2017) can compare different IB measures, this relies largely on between-study comparisons. However, when the SST has been compared to other IB measures within one study, similar results have often been found. For example, a recent study by O'Connor, Everaert, & Fitzgerald (2021) found that the SST showed the highest correlation with symptoms of depression compared to other measures of IBs including questionnaires and non-self-report measures, as well as the highest reliability. Further, the SST was the only measure to explain unique variance in depression symptoms, while its correlation with the other IB measures was weak to moderate. In another study, Rohrbacher (2016) found that both the convergent validity and reliability of the self-report Ambiguous Scenarios Test for Depression (AST-D; Rohrbacher & Reinecke, 2014) were slightly better than those of the SST, but that the AST and the SST were highly correlated. However, other non-questionnaire based measures showed only poor psychometric properties and were weakly correlated with the SST. In the context of subclinical panic disorder, Zahler et al. (2020) found that the interpretation bias questionnaire and the SST showed the most convincing psychometric properties of the IB measures administered and were highly correlated. Meanwhile other non-self-report IB measures demonstrated only weak convergent validity and small correlations with the SST. Finally, in a recent study in the context of Female Sexual Dysfunction (Zahler et al., 2021), the SST showed a higher correlation with sexual distress than an open ended scenario task to which it was only moderately correlated.

In summary, in the context of research on other IB measures, our findings indicate that the SST shows validity and reliability indices that are comparable to self-report measures, but there is evidence that its psychometric qualities are better than other measures that are not based on self-report. However, these conclusions are drawn largely from indirect comparisons across different studies, and the relationship between the SST and other IB measures will be important to investigate and synthesize more systematically in future research (see below).

4.6. Future directions

In the following, we present some perspectives for future research,

building on both the specific results of this review and broader considerations, including current challenges and opportunities. First, considering our promising findings for convergent validity and mixed findings for divergent validity, systematic validation studies are necessary, ideally including disorder-specific and disorder non-specific measures, and evaluating a range of psychometric indices. These studies could compare different SST versions to investigate which correlate most strongly with symptoms of the target disorder and most weakly with symptoms of other disorders. This could in turn inform subsequent modifications of the SST to enhance its specificity and increase divergent validity. A second and equally important consequence of such validation studies could be the development of 'gold standard' SSTs, within and across different disorder domains. Development of such standardized task versions (e.g., in terms of number of trials or stimuli) could help reduce the task's heterogeneity and enhance comparability across studies. In fact, it is likely that the high levels of heterogeneity in terms of SST versions used had a negative impact on the validity and reliability indices we found. Unfortunately, the results of our post-hoc moderation analyses did not help explain this heterogeneity, with the exception that computerized administration format was associated with higher convergent validity and reliability indices than other (e.g., pen-and-paper) formats. Accordingly, we would recommend that future studies use fully computerized versions of the SST. In fact, computerized versions of the SST offer a number of additional advantages. For example, via online administration of the SST larger and more diverse samples can be tested more easily over longer time-periods. Further, computerized versions allow the implementation of an exact time limit for each trial instead of defining a total time available for all stimuli, increasing the SST's standardization. While standardization of the SST and its administration would be helpful in enhancing comparability across studies, standardization is also important in relation to reporting. As previously mentioned, only a few studies reported details of SST performance such as the error rate and the cognitive load retention rate. However, this information is crucial for two reasons: First, while participants' performance on the SST itself and the correct recall of the cognitive load depend on many factors, both aspects are likely related to participants' engagement with the task and the task's difficulty. As such, task performance and correct recall provide valuable indicators of the results' validity. A second aspect relates to the number of error trials. The SST's analysis usually only includes correct trials, and it is therefore important to know how many trials the actual analysis included. This is particularly important from a reliability perspective, since reliability indices can fluctuate and can be strongly influenced by the number of trials. As such, we would encourage standardized reporting for the SST, including not only reliability indices but also information about the error and the cognitive load retention rates. We would also encourage researchers to make item-by-item individual participant response data on the SST available as this would allow other researchers to investigate response patterns or alternative scoring methods in more detail (e.g., mixed models with random intercepts at participant and stimulus levels, as is common in other areas of psychology, DeBruine & Barr, 2020).

Finally, although we investigated test-retest reliability in the current review, there was little data available; hence this is an area where much further research is warranted. Such research is necessary to estimate correlation indices between two SST assessments both within one laboratory session and over longer time-scales, to allow better interpretation of the effects of experimental manipulations and treatment interventions, and of longitudinal data. Optimally, this should be combined with investigation of how best to administer the SST repeatedly over time, since this poses a number of challenges (e.g., variations in stimuli).

While the above future directions arise directly from the results of our systematic review, consideration of the literature included raises some broader suggestions for future research. One of these concerns language and cross-cultural considerations. While there is preliminary evidence that the SST can be applied in different languages and alphabetical systems (e.g., in Chinese or Farsi), it is unclear to what extent the SST is comparable across different languages and cultures. Another consideration for future research is the question of which processes (on a continuum ranging from controlled to automatic IBs) the SST targets. Initially, the SST was designed to tap into more automatic processes when investigating IBs in comparison to self-report, namely by applying a cognitive load and a time limit (Wenzlaff & Bates, 1998). However, in some studies the SST has been considered to be a measure of more controlled IBs as participants are still required to evaluate the presented words (Everaert et al., 2017). Additionally, the SST has been considered to be a rather transparent measure in terms of what it assesses (cf. Phillips & Hine, 2013), which in turn could enable participants to control their answers. Our own meta-analytic results do not support the idea that the cognitive load alters the validity or reliability of the SST, at least in relation to cross-sectional data. However, there is data from longitudinal (e.g. Rude, Wenzlaff, Gibbs, Vane, & Whitney, 2002; Rude et al., 2003) and interventional (e.g., Bowler et al., 2012) studies indicating that the cognitive load may enable the SST to assess more automatic aspects of IBs, with prevention of thought suppression suggested as a mechanism (Geiger, Peters, & Baer, 2014). This would argue in favor of using a cognitive load when administering the SST, and against categorizing the SST under cognitive load as equivalent to pure self-report measures. In fact, it seems most plausible that if the SST indeed captures critical aspects of interpretational processing, this includes some aspects of both automatic and controlled processes.

While the current review has considered the SST as a behavioral measure of IBs in isolation, some particularly promising future directions concern how the basic SST may be combined with other methods. For example, the SST has been adapted so that eye-movement data can be tracked to assess participants' visual attention during the task (e.g., Sanchez, Everaert, De Putter, Mueller, & Koster, 2015; Sfärlea et al., 2020). This enables the simultaneous investigation of both attentional and interpretational biases. Given that the participant's choice of words is potentially influenced by their attentional processing of the words (e.g., fast engagement with certain types of words), it is also possible that the SST may reflect attention biases. This offers intriguing new research avenues, especially since it is unlikely that attentional and interpretation biases occur in isolation in psychopathology (Everaert, Tierens, Uzieblo, & Koster, 2013). More recently, the SST has also been tested in an fMRI setting (e.g., Viviani, Dommes, Bosch, Stingl, & Beschoner, 2018) by adapting the instructions so that participants are instructed to form the sentences in their mind and then press a button to indicate the sentence's valence instead of clicking words or writing them down. Given the promising results found for the SST by Viviani et al., the application of the SST in an fMRI setting could be a valuable tool to investigate the neural substrates of IBs, offering additional markers for interpretation biases.

4.7. Limitations of the review

This review is not without limitations. First, the interpretation of the convergent and divergent validity indices relied on the strength of the correlations between the SST and disorder-specific and -nonspecific symptom questionnaires, respectively. While we chose symptom questionnaires as the criterion against which the SST was validated (due to their well-established psychometric properties and common application), this operationalization of convergent validity assumes a strong theoretical relationship between IBs and symptoms. For disorders where there is a wealth of data supporting a relationship between other measures of IB and symptoms this may not in itself be problematic, but it makes it more difficult to compare the convergent validity indices across disorders; it may be that e.g., small correlations between the SST and symptoms of a particular disorder do not reflect poor validity of the measure, but rather that IBs do not play such a central role in that disorder. Ultimately, questions about the role or importance of IBs in particular disorders cannot be answered via cross-sectional studies, but

rather require longitudinal and interventional designs, and this is a crucial area for future research. However, this consideration also highlights the potential utility of including multiple IB measures in single studies. As a second limitation, the validity and reliability indices were summarized across different sample types such as patients, healthy college students, adolescents etc., which may have introduced heterogeneity to the results. Despite our attempt to explain the between-study heterogeneity through moderation analyses, the large amount of unaccounted heterogeneity poses a clear limitation to the interpretability of our meta-analytical findings. In relation to these moderation analyses, it should be noted that the moderators were chosen post-hoc, based on face-validity given the context of the present review and meta-analysis. Finally, our Risk of Bias Assessment (ROB) comes with some caveats: Although we used an established measure as the basis for assessing ROB we had to make some adaptations given the specific aims of this review. Further, this was not the original ROB process specified in our pre-registration but rather adopted as an improvement upon this during the review process.

5. Conclusion

The SST is a widely applied measure of IBs across a broad range of samples and disorders. Overall, convergent validity and internal consistency are promising, with more mixed findings for divergent validity. While the SST offer promising routes for future research, variability between studies highlights the importance of standardization of both the SST itself and the reporting of its results.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpr.2022.102133.

Contributors

Felix Würtz: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – Original draft, Writing – Review and editing, Visualization.

Lisa Zahler: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – Original draft, Writing – Review and editing.

Simon E. Blackwell: Conceptualization, Methodology, Investigation, Writing – Review and editing, Supervision, Project administration.

Mahdi Bagheri: Investigation, Validation, Writing – Review and editing.

Jürgen Margraf: Conceptualization, Methodology, Writing – Review and editing, Supervision, Project administration.

Marcella L. Woud: Conceptualization, Methodology, Investigation, Writing – Review and editing, Supervision, Project administration.

Declaration of Competing Interest

None.

Appendix A. References used in the qualitative/quantitative review

Beevers, C. G. (2002). *Cognitive bias and affect regulation as prospective predictors of depressive symptoms.* Unpublished Doctoral Thesis, University of Miami, Miami, United States.

Beevers, C. G., & Meyer, B. (2008). I feel fine but the glass is still half empty: Thought suppression biases information processing despite recovery from a dysphoric mood state. *Cognitive Therapy and Research, 32* (3), 323–332. https://doi.org/10.1007/s10608-006-9108-8

Bibi, A., Margraf, J., & Blackwell, S. E. (2020). Positive imagery cognitive bias modification for symptoms of depression among university students in Pakistan: A pilot study. *Journal of Experimental Psychopathology, 11*(2), 1–22. https://doi.org/10.1177/2043808720918030

Blackwell, S. E., Browning, M., Mathews, A., Pictet, A., Welch, J., Davies, J., … Holmes, E. A. (2015). Positive imagery-based cognitive bias modification as a web-based treatment tool for depressed adults: A randomized controlled trial. *Clinical Psychological Science, 3*(1), 91–111. https://doi.org/10.1177/2167702614560746

Blackwell, S. E., & Holmes, E. A. (2010). Modifying Interpretation and Imagination in Clinical Depression: A Single Case Series Using Cognitive Bias Modification. *Applied Cognitive Psychology, 24*, 338–350. https://doi.org/10.1002/acp.1680

Blackwell, S. E., & Holmes, E. A. (2017). Brightening the day with flashes of positive mental imagery: A case study of an individual with depression. *Journal of Clinical Psychology, 73*(5), 579–589. https://doi.org/10.1002/jclp.22455

Blanco, I., Boemo, T., & Sanchez-Lopez, A. (2021). The role of cognitive biases and emotion regulation strategies when facing major stressors: ecological evidence during the COVID-19 lockdown of 2020 using a novel online cognitive bias assessment. *JMIR Preprints.* https://doi.org/10.2196/preprints.30961

Bowler, J. O., MacKintosh, B., Dunn, B. D., Mathews, A., Dalgleish, T., & Hoppitt, L. (2012). A comparison of cognitive bias modification for interpretation and computerized cognitive behavior therapy: Effects on anxiety, depression, attentional control, and interpretive bias. *Journal of Consulting and Clinical Psychology, 80*(6), 1021–1033. https://doi.org/10.1037/a0029932

Bradatsch, S., Vahl, M. D., Potterton, R., Gordon, G., Schmidt, U., & Brockmeyer, T. (2020). Interpretation bias modification to reduce body dissatisfaction – a randomized controlled pilot study in women with elevated weight and shape concerns. *Journal of Eating Disorders, 8*(1), 34. https://doi.org/10.1186/s40337-020-00305-4

Brockmeyer, T., Anderle, A., Schmidt, H., Febry, S., Wünsch-Leiteritz, W., Leiteritz, A., & Friederich, H. C. (2018). Body image related negative interpretation bias in anorexia nervosa. *Behaviour Research and Therapy, 104*, 69–73. https://doi.org/10.1016/j.brat.2018.03.003

Brockmeyer, T., Pfeiffer, N., Grosse Holtforth, M., Zimmermann, J., Kämmerer, A., Friederich, H.-C., & Bents, H. (2012). Mood Regulation and Cognitive Reactivity in Depression Vulnerability. *Cognitive Therapy and Research, 36*(6), 634–642. https://doi.org/10.1007/s10608-011-9406-7

Burnett Heyes, S., Pictet, A., Mitchell, H., Raeder, S. M., Lau, J. Y. F., Holmes, E. A., & Blackwell, S. E. (2017). Mental Imagery-Based Training to Modify Mood and Cognitive Bias in Adolescents: Effects of Valence and Perspective. *Cognitive Therapy and Research, 41*(1), 73–88. https://doi.org/10.1007/s10608-016-9795-8

Cowden Hindash, A. H., & Rottenberg, J. (2017a). Turning quickly on myself: Automatic interpretation biases in dysphoria are self-referent. *Cognition and Emotion*, *31*(2), 395–402. https://doi.org/10.1080/02699931.2015.1105792

Unpublished data collected during Cowden Hindash, A. H., & Rottenberg, J. A. (2017b). Moving towards the benign: Automatic interpretation bias modification in dysphoria. *Behaviour Research and Therapy*, *99*, 98–107. https://doi.org/10.1016/j.brat.2017.09.005

de Voogd, E. L., De Hullu, E., Heyes, S. B., Blackwell, S. E., Wiers, R. W., & Salemink, E. (2017). Imagine the bright side of life: A randomized controlled trial of two types of interpretation bias modification procedure targeting adolescent anxiety and depression. *PLoS ONE*, *12*(7), 1–24. https://doi.org/10.1371/journal.pone.0181147

Deschênes, S. S., Dugas, M. J., Anderson, K. G., & Gouin, J.-P. (2015). Effects of anger on interpretation bias, negative beliefs about uncertainty, and worry catastrophizing: An experimental study. *Journal of Experimental Psychopathology*, *6*(2), 138–148. https://doi.org/10.5127/jep.042814

Dickson, D. A. (2015). *Investigating the role of cognitive biases as a risk factor for depression* (Doctoral Thesis, Loyola University Chicago, Chicago, United States). Doctoral Thesis, Loyola University Chicago, Chicago, United States. Retrieved from https://ecommons.luc.edu/luc_diss/1471/?utm_source=ecommons.luc.edu%2Fluc_diss%2F1471&utm_medium=PDF&utm_campaign=PDFCoverPages

Donegan, E. (2016). *CBT and Interpretation Bias Modification for Generalized Anxiety Disorder: Examining the Roles of Intolerance of Uncertainty and Interpretation Bias in Symptom Reduction* (Doctoral Thesis, Concordia Unversity, Montréal, Québec, Canada). Doctoral Thesis, Concordia Unversity, Montréal, Québec, Canada. Retrieved from https://spectrum.library.concordia.ca/982539/

Everaert, J., Duyck, W., & Koster, E. H. W. (2014). Attention, interpretation, and memory biases in subclinical depression: A proof-of-principle test of the combined cognitive biases hypothesis. *Emotion*, *14*(2), 331–340. https://doi.org/10.1037/a0035250

Everaert, J., Grahek, I., Duyck, W., Buelens, J., Van den Bergh, N., & Koster, E. H. W. (2017). Mapping the interplay among cognitive biases, emotion regulation, and depressive symptoms. *Cognition & Emotion*, *31*(4), 726–735. https://doi.org/10.1080/02699931.2016.1144561

Everaert, J., Grahek, I., & Koster, E. H. W. (2016). Individual differences in cognitive control over emotional material modulate cognitive biases linked to depressive symptoms. *Cognition and Emotion*, *31*(4), 736–746. https://doi.org/10.1080/02699931.2016.1144562

Everaert, J., Mogoașe, C., David, D., & Koster, E. H. W. W. (2015). Attention bias modification via single-session dot-probe training: Failures to replicate. *Journal of Behavior Therapy and Experimental Psychiatry*, *49*(Part A), 5–12. https://doi.org/10.1016/j.jbtep.2014.10.011

Everaert, J., Tierens, M., Uzieblo, K., & Koster, E. H. W. (2013). The indirect effect of attention bias on memory via interpretation bias: evidence for the combined cognitive bias hypothesis in subclinical depression. *Cognition & Emotion*, *27*(8), 1450–1459. https://doi.org/10.1080/02699931.2013.787972

Geiger, P. J., Peters, J. R., & Baer, R. A. (2014). Using a measure of cognitive distortion to examine the relationship between thought suppression and borderline personality features: A multi-method investigation. *Personality and Individual Differences*, *59*, 54–59. https://doi.org/10.1016/j.paid.2013.11.005

Gill, S. E., Miller, J. V., Haner, M. L., & Rude, S. S. (2017). Putting it into context: Creating a self-report measure of Big Picture Appraisal. *International Journal of Cognitive Therapy*, *10*(3), 234–253. https://doi.org/10.1521/ijct.2017.10.3.234

Hedlund, S., & Rude, S. S. (1995). Evidence of Latent Depressive Schemas in Formerly Depressed Individuals. *Journal of Abnormal Psychology*, *104*(3), 517–525. https://doi.org/10.1037/0021-843X.104.3.517

Hirsch, C. R., Krahé, C., Whyte, J., Bridge, L., Loizou, S., Norton, S., & Mathews, A. (2020). Effects of modifying interpretation bias on transdiagnostic repetitive negative thinking. *Journal of Consulting and Clinical Psychology*, *88*(3), 226–239. https://doi.org/10.1037/ccp0000455

Hirsch, C. R., Krahé, C., Whyte, J., Krzyzanowski, H., Meeten, F., Norton, S., & Mathews, A. (2021). Internet-delivered interpretation training reduces worry and anxiety in individuals with generalized anxiety disorder: A randomized controlled experiment. *Journal of Consulting and Clinical Psychology*, *89*(7), 575–589. https://doi.org/10.1037/ccp0000660

Hirsch, C. R., Krahé, C., Whyte, J., Loizou, S., Bridge, L., Norton, S., & Mathews, A. (2018). Interpretation training to target repetitive negative thinking in generalized anxiety disorder and depression. *Journal of Consulting and Clinical Psychology*, *86*(12), 1017–1030. https://doi.org/10.1037/ccp0000310

Hirsch, C. R., Meeten, F., Gordon, C., Newby, J. M., Bick, D., & Moulds, M. L. (2020). Repetitive negative thinking and interpretation bias in pregnancy. *Clinical Psychology in Europe*, *2*(4). https://doi.org/10.32872/cpe.v2i4.3615

Holmes, E. A., Lang, T. J., & Shah, D. M. (2009). Developing Interpretation Bias Modification as a "Cognitive Vaccine" for Depressed Mood: Imagining Positive Events Makes You Feel Better Than Thinking About Them Verbally. *Journal of Abnormal Psychology*, *118*(1), 76–88. https://doi.org/10.1037/a0012590

Ji, J. L., Holmes, E. A., & Blackwell, S. E. (2017). Seeing light at the end of the tunnel: Positive prospective mental imagery and optimism in depression. *Psychiatry Research*, *247*, 155–162. https://doi.org/10.1016/j.psychres.2016.11.025

Krahé, C., Meeten, F., & Hirsch, C. R. (2020). *Development and Validation of the Scrambled Sentences Test for Worry*.

Krahé, C., Whyte, J., Bridge, L., Loizou, S., & Hirsch, C. R. (2019). Are different forms of repetitive negative thinking associated with interpretation bias in generalized anxiety disorder and depression? *Clinical Psychological Science*, *7*(5), 969–981. https://doi.org/10.1177/2167702619851808

Krans, J., Bosmans, G., Salemink, E., & De Raedt, R. (2019). Cognitive Bias Modification of expectancies (CBM-E): Effects on interpretation bias and autobiographical memory, and relations with social and attachment anxiety. *Cognitive Therapy and Research*, *43*(6), 1028–1042. https://doi.org/10.1007/s10608-019-10032-z

Krejtz, I., Holas, P., Rusanowska, M., & Nezlek, J. B. (2018). Positive online attentional training as a means of modifying attentional and interpretational biases among the clinically depressed: An experimental study using eye tracking. *Journal of Clinical Psychology*, *74*(9), 1594–1606. https://doi.org/10.1002/jclp.22617

Lang, T. J., Blackwell, S. E., Harmer, C. J., Davison, P., & Holmes, E. A. (2012). Cognitive Bias Modification Using Mental Imagery for Depression: Developing a Novel Computerized Intervention to Change Negative Thinking Styles. *European Journal of Personality*, *26*(2), 145–157. https://doi.org/10.1002/per.855

Lee, J. S., Mathews, A., Shergill, S., & Yiend, J. (2016). Magnitude of negative interpretation bias depends on severity of depression. *Behaviour Research and Therapy*, *83*, 26–34. https://doi.org/10.1016/j.brat.2016.05.007

LeMoult, J., Colich, N., Joormann, J., Singh, M. K., Eggleston, C., & Gotlib, I. H. (2018). Interpretation bias training in depressed adolescents: Near- and far-transfer effects. *Journal of Abnormal Child Psychology*, *46*(1), 159–167. https://doi.org/10.1007/s10802-017-0285-6

Lin, X.-X., Si, S.-W., Gao, R.-R., Sun, Y.-B., Wang, Y.-Z., Wang, N., … Wang, J.-Y. (2020). Does Approach-Avoidance Behavior in Response to Ambiguous Cues Reflect Depressive Interpretation Bias? Related but Distinct. *Cognitive Therapy and Research*, *44*(6), 1091–1105. https://doi.org/10.1007/s10608-020-10133-0

Lo, H. (2009). *Thought suppression as a cognitive vulnerability factor for depression – an fMRI study* (Doctoral Thesis, Ulm University, Ulm, Germany). Doctoral Thesis, Ulm University, Ulm, Germany. Retrieved from https://oparu.uni-ulm.de/xmlui/handle/123456789/2125

Maestas, K. L., Amidon, A., Baum, E. S., Chrisman, J. G., Durham, J. A., Rooney, S. B., … Swann, W. B. (2008). Partner devaluation is associated with depression symptoms among depression-vulnerable women with low self-esteem. *Journal of Social and Clinical Psychology*, *27*(6), 621–640. https://doi.org/10.1521/jscp.2008.27.6.621

Marchetti, I., Koster, E. H. W., & De Raedt, R. (2012). Mind-wandering heightens the accessibility of negative relative to positive thought. *Consciousness and Cognition*, *21*(3), 1517–1525. https://doi.org/10.1016/j.concog.2012.05.013

Martin-Romero, N., Sanchez-Lopez, A., & Vazquez, C. (in prep). *Interpretation Bias as a Clinical and Vulnerability Marker for Depression: New Insights Derived From the Scrambled Sentence Task.* Manuscript in preparation.

McNally, L. (2014). *A Cognitive Bias Modification for Interpretation (CBM-I) Task with Individuals Experiencing Clinical Levels of Generalised Anxiety: A Single Case Series* (Doctoral Thesis, University of East Anglia, Norwich, Norfolk). Doctoral Thesis, University of East Anglia, Norwich, Norfolk. Retrieved from https://ueaeprints.uea.ac.uk/id/eprint/50726/

Miklowitz, D. J., Alatiq, Y., Geddes, J. R., Goodwin, G. M., & Williams, J. M. G. (2010). Thought suppression in patients with bipolar disorder. *Journal of Abnormal Psychology*, *119*(2), 355–365. https://doi.org/10.1037/a0018613

Miller, J. V., Rude, S. S., & Haner, M. (2015). Experimental Modification of Appraisal Style: Benefits of Seeing the Big Picture. *Journal of Experimental Psychopathology*, *6*(1), 100–111. https://doi.org/10.5127/jep.041914

Moon, S. M., & Lord, R. G. (2006). Individual differences in automatic and controlled regulation of emotion and task performance. *Human Performance*, *19*(4), 327–356. https://doi.org/10.1207/s15327043hup1904_2

Murphy, S. E., O'Donoghue, M. C., Drazich, E. H. S., Blackwell, S. E., Nobre, A. C., & Holmes, E. A. (2015). Imagining a brighter future: The effect of positive imagery training on mood, prospective mental imagery and emotional bias in older adults. *Psychiatry Research*, *230*(1), 36–43. https://doi.org/10.1016/j.psychres.2015.07.059

Novović, Z., Mihić, L., Biro, M., & Tovilović, S. (2014). Measuring vulnerability to depression: The serbian scrambled sentences test-SSST. *Psihologija*, *47*(1), 33–48. https://doi.org/10.2298/PSI1401033N

O'Connor, C. E., Everaert, J., & Fitzgerald, A. (2021). Interpreting ambiguous emotional information: Convergence among interpretation bias measures and unique relations with depression severity. *Journal of Clinical Psychology*, (April), 1–16. https://doi.org/10.1002/jclp.23186

Penton-Voak, I. S., Adams, S., Button, K. S., Fluharty, M., Dalili, M., Browning, M., … Munafò, M. R. (2020). Emotional recognition training modifies neural response to emotional faces but does not improve mood in healthy volunteers with high levels of depressive symptoms. *Psychological Medicine*. https://doi.org/10.1017/S0033291719004124

Phillips, W. J., & Hine, D. W. (2013). Exploring the Factor Structure of Implicit and Explicit Cognitions Associated With Depression. *Assessment*, *20*(4), 474–483. https://doi.org/10.1177/1073191112437595

Phillips, W. J., Hine, D. W., & Bhullar, N. (2012). A latent profile analysis of implicit and explicit cognitions associated with depression. *Cognitive Therapy and Research*, *36*(5), 458–473. https://doi.org/10.1007/s10608-011-9381-z

Romero, N., Sanchez, A., & Vazquez, C. (2014). Memory biases in remitted depression: the role of negative cognitions at explicit and automatic processing levels. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*(1), 128–135. https://doi.org/10.1016/j.jbtep.2013.09.008

Rude, S. S., Covich, J., Jarrold, W., Hedlund, S., & Zentner, M. (2001). Detecting Depressive Schemata in Vulnerable Individuals: Questionnaires Versus Laboratory Tasks. *Cognitive Therapy and Research*, *25*(1), 103–116. https://doi.org/10.1023/A:1026482917201

Rude, S. S., Durham-Fowler, J. A., Baum, E. S., Rooney, S. B., & Maestas, K. L. (2010). Self-report and cognitive processing measures of depressive thinking predict subsequent major depressive disorder.

Cognitive Therapy and Research, *34*(2), 107–115. https://doi.org/10.1007/s10608-009-9237-y

Unpublished data collected during Rude, S. S., Durham-Fowler, J. A., Baum, E. S., Rooney, S. B., & Maestas, K. L. (2010). Self-report and cognitive processing measures of depressive thinking predict subsequent major depressive disorder. *Cognitive Therapy and Research*, *34*(2), 107–115. https://doi.org/10.1007/s10608-009-9237-y

Rude, S. S., Mazzetti, F. A., Pal, H., & Stauble, M. R. (2011). Social rejection: How best to think about it? *Cognitive Therapy and Research*, *35*(3), 209–216. https://doi.org/10.1007/s10608-010-9296-0

Rude, S. S., Valdez, C. R., Odom, S., & Ebrahimi, A. (2003). Negative cognitive biases predict subsequent depression. *Cognitive Therapy and Research*, *27*(4), 415–429. https://doi.org/10.1023/A:1025472413805

Rude, S. S., Wenzlaff, R. M., Gibbs, B., Vane, J., & Whitney, T. (2002). Negative processing biases predict subsequent depressive symptoms. *Cognition and Emotion*, *16*(3), 423–440. https://doi.org/10.1080/02699930143000554

Sanchez-Lopez, A., De Raedt, R., van Put, J., & Koster, E. H. W. (2019). A novel process-based approach to improve resilience: Effects of computerized mouse-based (gaze)contingent attention training (MCAT) on reappraisal and rumination. *Behaviour Research and Therapy*, *118* (April), 110–120. https://doi.org/10.1016/j.brat.2019.04.005

Sanchez-Lopez, A., Everaert, J., Van Put, J., De Raedt, R., & Koster, E. H. W. (2019). Eye-gaze contingent attention training (ECAT): Examining the causal role of attention regulation in reappraisal and rumination. *Biological Psychology*, *142*(January), 116–125. https://doi.org/10.1016/j.biopsycho.2019.01.017

Sanchez, A., Duque, A., Romero, N., & Vazquez, C. (2017). Disentangling the Interplay Among Cognitive Biases: Evidence of Combined Effects of Attention, Interpretation and Autobiographical Memory in Depression. *Cognitive Therapy and Research*, *41*(6), 829–841. https://doi.org/10.1007/s10608-017-9858-5

Sanchez, A., Everaert, J., De Putter, L. M. S., Mueller, S. C., & Koster, E. H. W. (2015). Life is … great! Emotional attention during instructed and uninstructed ambiguity resolution in relation to depressive symptoms. *Biological Psychology*, *109*, 67–72. https://doi.org/10.1016/j.biopsycho.2015.04.007

Sanchez, A., Everaert, J., & Koster, E. H. W. (2016). Attention training through gaze-contingent feedback: Effects on reappraisal and negative emotions. *Emotion*, *16*(7), 1074–1085. https://doi.org/10.1037/emo0000198

Savulich, G., Edwards, A., Assadi, S., Guven, H., Leathers-Smith, E., Shergill, S., & Yiend, J. (2020). Biased interpretation in paranoia and its modification. *Journal of Behavior Therapy and Experimental Psychiatry*, *69* (May), 101575. https://doi.org/10.1016/j.jbtep.2020.101575

Savulich, G., Freeman, D., Shergill, S., & Yiend, J. (2015). Interpretation biases in paranoia. *Behavior Therapy*, *46*(1), 110–124. https://doi.org/10.1016/j.beth.2014.08.002

Savulich, G., Shergill, S. S., & Yiend, J. (2017). Interpretation biases in clinical paranoia. *Clinical Psychological Science*, *5*(6), 985–1000. https://doi.org/10.1177/2167702617718180

Seeds, P. M. (2012). *Interpretive Bias in the Context of Life Stress and Depression: An Examination of Stress Generation and Diathesis-Stress Models* (Doctoral Thesis, University of Western Ontario, London, Ontario, Canada). Doctoral Thesis, University of Western Ontario, London, Ontario, Canada. Retrieved from https://ir.lib.uwo.ca/etd/664

Sfärlea, A, Takano, K., Buhl, C., Loechner, J., Greimel, E., Salemink, E., … Platt, B. (2021). Emotion Regulation as a Mediator in the Relationship Between Cognitive Biases and Depressive Symptoms in Depressed, At-risk and Healthy Children and Adolescents. *Research on Child and Adolescent Psychopathology*, *49*(10), 1345–1358. https://doi.org/10.1007/s10802-021-00814-z

Sfärlea, Anca, Buhl, C., Loechner, J., Neumüller, J., Asperud Thomsen, L., Starman, K., … Platt, B. (2020). "I Am a Total…Loser" - The Role of Interpretation Biases in Youth Depression. *Journal of Abnormal Child Psychology*, *48*(10), 1337–1350. https://doi.org/10.1007/s10802-020-

00670-3

Sfärlea, Anca, Löchner, J., Neumüller, J., Asperud Thomsen, L., Starman, K., Salemink, E., … Platt, B. (2019). Passing on the half-empty glass: A transgenerational study of interpretation biases in children at risk for depression and their parents with depression. *Journal of Abnormal Psychology*, *128*(2), 151–161. https://doi.org/10.1037/abn0000401

Standage, H., Ashwin, C., & Fox, E. (2010). Is manipulation of mood a critical component of cognitive bias modification procedures? *Behaviour Research and Therapy*, *48*(1), 4–10. https://doi.org/10.1016/j.brat.2009.08.005

Torkan, H., Blackwell, S. E., Holmes, E. A., Kalantari, M., Neshat-Doost, H. T., Maroufi, M., & Talebi, H. (2014). Positive imagery cognitive bias modification in treatment-seeking patients with major depression in Iran: A pilot study. *Cognitive Therapy and Research*, *38*(2), 132–145. https://doi.org/10.1007/s10608-014-9598-8

Van der Does, A. J. W., Manthey, L., & Hermans, K. (2012). Dysphoric mood, thought suppression and cognitive control in recovered depressed and never-depressed individuals. *Journal of Experimental Psychopathology*, *3*(1), 69–83. https://doi.org/10.5127/jep.20711

Van der Does, W. (2005). Thought suppression and cognitive vulnerability to depression. *The British Journal of Clinical Psychology*, *44* (Pt 1), 1–14. https://doi.org/10.1348/014466504x19442

Viviani, R., Dommes, L., Bosch, J. E., Stingl, J. C., & Beschoner, P. (2018). A Computerized Version of the Scrambled Sentences Test. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02310

Viviani, R., Lo, H., Sim, E. J., Beschoner, P., Stingl, J. C., & Horn, A. B. (2010). The neural substrate of positive bias in spontaneous emotional processing. *PLoS ONE*, *5*(11), 1–9. https://doi.org/10.1371/journal.pone.0015454

Viviani, R., Mahler, J. S., Bosch, J. E., Dommes, L., Eberhardt, J., Beschoner, P., & Sosic-Vasic, Z. (2018). Adaptation of the Scrambled-Sentences Task to Assess "Shattered Assumptions:" Construction of the Test and Investigation of Neural Substrates in an fMRI Study. *Psychopathology*, *51*(5), 306–317. https://doi.org/10.1159/000491701

Watkins, E. R., & Moulds, M. (2007). Revealing negative thinking in recovered major depression: A preliminary investigation. *Behaviour Research and Therapy*, *45*(12), 3069–3076. https://doi.org/10.1016/j.brat.2007.05.001

Wells, T. T., Vanderlind, W. M., Selby, E. A., & Beevers, C. G. (2014). Childhood abuse and vulnerability to depression: Cognitive scars in otherwise healthy young adults. *Cognition and Emotion*, *28*(5), 821–833. https://doi.org/10.1080/02699931.2013.864258

Wenzlaff, R. M., & Bates, D. E. (1998). Unmasking a cognitive vulnerability to depression: how lapses in mental control reveal depressive thinking. *Journal of Personality and Social Psychology*, *75*(6), 1559–1571. https://doi.org/10.1037/0022-3514.75.6.1559

Westermann, K., Woud, M. L., Cwik, J. C., Graz, C., Nyhuis, P., Margraf, J., & Blackwell, S. E. (2021). Feasibility of computerised positive mental imagery training as a treatment adjunct in in-patient mental health settings: Randomised controlled trial. *BJPsych Open, 7*(6), E203. doi:10.1192/bjo.2021.1042

Williams, A. D., Blackwell, S. E., Mackenzie, A., Holmes, E. A., & Andrews, G. (2013). Combining imagination and reason in the treatment of depression: A randomized controlled trial of internet-based cognitive-bias modification and internet-CBT for depression. *Journal of Consulting and Clinical Psychology*, *81*(5), 793–799. https://doi.org/10.1037/a0033247

Würtz, F., Blackwell, S. E., Margraf, J., & Woud, M. L. (2021). *The World Dangerous it is – The Scrambled Sentences Task for Negative Life Events*. Preprint available at: https://doi.org/10.31234/osf.io/hgxd4

Yiend, J., Allen, P., Lopez, N. D., Falkenberg, I., Tseng, H.-H., & McGuire, P. (2019). Negative Interpretation Biases Precede the Onset of Psychosis. *Behavior Therapy*, *50*(4), 718–731. https://doi.org/10.1016/j.beth.2018.11.002

Yiend, J., Lee, J. S., Tekes, S., Atkins, L., Mathews, A., Vrinten, M., …

Shergill, S. (2014). Modifying interpretation in a clinically depressed sample using "cognitive bias modification-errors": A double blind randomised controlled trial. *Cognitive Therapy and Research*, *38*(2), 146–159. https://doi.org/10.1007/s10608-013-9571-y

Zahler, L., Meyers, M., Woud, M. L., Blackwell, S. E., Margraf, J., & Velten, J. (2021). Using Three Indirect Measures to Assess the Role of Sexuality-Related Associations and Interpretations for Women's Sexual Desire: An Internet-Based Experimental Study. *Archives of Sexual Behavior*. https://doi.org/10.1007/s10508-020-01897-3

Zahler, L., Sommer, K., Reinecke, A., Wilhelm, F. H., Margraf, J., & Woud, M. L. (2020). Cognitive vulnerability in the context of panic: Assessment of panic-related associations and interpretations in individuals with varying levels of anxiety sensitivity. *Cognitive Therapy and Research*, *44*(4), 858–873. https://doi.org/10.1007/s10608-020-10103-6

References

Assink, M., & Wibbelink, C. (2016). Fitting three-level meta-analytic models in R : A step-by-step tutorial. *The Quantitative Methods for Psychology, 12*(3), 154–174.

Beard, C., & Amir, N. (2009). Interpretation in social anxiety: When meaning precedes ambiguity. *Cognitive Therapy and Research, 33*(4), 406–415. https://doi.org/10.1007/s10608-009-9235-0

Beck, A. T. (1976). *Cognitive therapy and the emotional disorders* (2nd ed.). International Universities Press.

Bowler, J. O., MacKintosh, B., Dunn, B. D., Mathews, A., Dalgleish, T., & Hoppitt, L. (2012). A comparison of cognitive bias modification for interpretation and computerized cognitive behavior therapy: Effects on anxiety, depression, attentional control, and interpretive bias. *Journal of Consulting and Clinical Psychology, 80*(6), 1021–1033. https://doi.org/10.1037/a0029932

Brockmeyer, T., Anderle, A., Schmidt, H., Febry, S., Wünsch-Leiteritz, W., Leiteritz, A., & Friederich, H. C. (2018). Body image related negative interpretation bias in anorexia nervosa. *Behaviour Research and Therapy, 104*, 69–73. https://doi.org/10.1016/j.brat.2018.03.003

Clark, D. M. (1986). A cognitive approach to panic. *Behaviour Research and Therapy, 24* (4), 461–470. https://doi.org/10.1016/0005-7967(86)90011-2

DeBruine, L. M., & Barr, D. J. (2020). Understanding mixed effects models through data simulation. *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.1177/2515245920965119

Everaert, J., Podina, I. R., & Koster, E. H. W. (2017). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review, 58*(September), 33–48. https://doi.org/10.1016/j.cpr.2017.09.005

Everaert, J., Tierens, M., Uzieblo, K., & Koster, E. H. W. (2013). The indirect effect of attention bias on memory via interpretation bias: Evidence for the combined cognitive bias hypothesis in subclinical depression. *Cognition and Emotion, 27*(8), 1450–1459. https://doi.org/10.1080/02699931.2013.787972

Fodor, L. A., Georgescu, R., Cuijpers, P., Szamoskozi, Ş., David, D., Furukawa, T. A., & Cristea, I. A. (2020). Efficacy of cognitive bias modification interventions in anxiety and depressive disorders: A systematic review and network meta-analysis. *The Lancet Psychiatry, 7*(6), 506–514. https://doi.org/10.1016/S2215-0366(20)30130-9

Geiger, P. J., Peters, J. R., & Baer, R. A. (2014). Using a measure of cognitive distortion to examine the relationship between thought suppression and borderline personality features: A multi-method investigation. *Personality and Individual Differences, 59*, 54–59. https://doi.org/10.1016/j.paid.2013.11.005

Gonsalves, M., Whittles, R. L., Weisberg, R. B., & Beard, C. (2019). A systematic review of the word sentence association paradigm (WSAP). *Journal of Behavior Therapy and Experimental Psychiatry, 64*, 133–148. https://doi.org/10.1016/j.jbtep.2019.04.003

Hirsch, C. R., Krahé, C., Whyte, J., Loizou, S., Bridge, L., Norton, S., & Mathews, A. (2018). Interpretation training to target repetitive negative thinking in generalized anxiety disorder and depression. *Journal of Consulting and Clinical Psychology, 86*(12), 1017–1030. https://doi.org/10.1037/ccp0000310

Hirsch, C. R., Meeten, F., Krahé, C., & Reeder, C. (2016). Resolving ambiguity in emotional disorders: The nature and role of interpretation biases. *Annual Review of Clinical Psychology, 12*, 281–305. https://doi.org/10.1146/annurev-clinpsy-021815-093436

Jones, E. B., & Sharpe, L. (2017). Cognitive bias modification: A review of meta-analyses. *Journal of Affective Disorders, 223*, 175–183. https://doi.org/10.1016/j.jad.2017.07.034

Joormann, J., Waugh, C. E., & Gotlib, I. H. (2015). Cognitive bias modification for interpretation in major depression: Effects on memory and stress reactivity. *Clinical Psychological Science, 3*(1), 126–139. https://doi.org/10.1177/2167702614560748

Kleim, B., Grey, N., Wild, J., Nussbeck, F. W., Stott, R., Hackmann, A., Clark, D. M., & Ehlers, A. (2013). Cognitive change predicts symptom reduction with cognitive therapy for posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 81*(3), 383–393. https://doi.org/10.1037/a0031290

Knowles, K. A., & Olatunji, B. O. (2020). Specificity of trait anxiety in anxiety and depression: Meta-analysis of the state-trait anxiety inventory. *Clinical Psychology Review, 82*, Article 101928. https://doi.org/10.1016/j.cpr.2020.101928

Korn, J., Dietel, F. A., & Hartmann, A. S. (2020). Testing the specificity of interpretation biases in women with eating disorder symptoms: An online experimental assessment.

International Journal of Eating Disorders, 53(3), 372–382. https://doi.org/10.1002/eat.23201

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54*(4), 337. https://doi.org/10.1001/archpsyc.1997.01830160065009

Krahé, C., Whyte, J., Bridge, L., Loizou, S., & Hirsch, C. R. (2019). Are different forms of repetitive negative thinking associated with interpretation bias in generalized anxiety disorder and depression? *Clinical Psychological Science, 7*(5), 969–981. https://doi.org/10.1177/2167702619851808

Lee, J. S., Mathews, A., Shergill, S., & Yiend, J. (2016). Magnitude of negative interpretation bias depends on severity of depression. *Behaviour Research and Therapy, 83*, 26–34. https://doi.org/10.1016/j.brat.2016.05.007

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Journal of Clinical Epidemiology, 62*(10), e1–e34. https://doi.org/10.1016/j.jclinepi.2009.06.006

Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology, 1*(1), 167–195. https://doi.org/10.1146/annurev.clinpsy.1.102803.143916

Miklowitz, D. J., Alatiq, Y., Geddes, J. R., Goodwin, G. M., & Williams, J. M. G. (2010). Thought suppression in patients with bipolar disorder. *Journal of Abnormal Psychology, 119*(2), 355–365. https://doi.org/10.1037/a0018613

National Heart Lung and Blood Institute. (2021). *Study Quality Assessment Tools*. https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools.

O'Connor, C. E., Everaert, J., & Fitzgerald, A. (2021). Interpreting ambiguous emotional information: Convergence among interpretation bias measures and unique relations with depression severity. *J Clin Psychol, 77*, 2529–2544. https://doi.org/10.1002/jclp.23186

Parsons, S., Kruijt, A.-W., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in Methods and Practices in Psychological Science, 2*(4), 378–395. https://doi.org/10.1177/2515245919879695

Phillips, W. J., & Hine, D. W. (2013). Exploring the factor structure of implicit and explicit cognitions associated with depression. *Assessment, 20*(4), 474–483. https://doi.org/10.1177/1073191112437595

Phillips, W. J., Hine, D. W., & Thorsteinsson, E. B. (2010). Implicit cognition and depression: A meta-analysis. *Clinical Psychology Review, 30*(6), 691–709. https://doi.org/10.1016/j.cpr.2010.05.002

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research (R package version 1.9.12)*. Northwestern University. https://cran.r-project.org/package=psych.

Rohrbacher, H. (2016). Interpretation bias in the context of depressed mood: Assessment strategies and the role of self-generation in cognitive bias modification. http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-207298.

Rohrbacher, H., & Reinecke, A. (2014). Measuring change in depression-related interpretation bias: Development and validation of a parallel ambiguous scenarios test. *Cognitive Behaviour Therapy, 43*(3), 239–250. https://doi.org/10.1080/16506073.2014.919605

Rude, S. S., Valdez, C. R., Odom, S., & Ebrahimi, A. (2003). Negative cognitive biases predict subsequent depression. *Cognitive Therapy and Research, 27*(4), 415–429. https://doi.org/10.1023/A:1025472413805

Rude, S. S., Wenzlaff, R. M., Gibbs, B., Vane, J., & Whitney, T. (2002). Negative processing biases predict subsequent depressive symptoms. *Cognition and Emotion, 16*(3), 423–440. https://doi.org/10.1080/02699930143000554

Sanchez, A., Everaert, J., De Putter, L. M. S., Mueller, S. C., & Koster, E. H. W. (2015). Life is … great! Emotional attention during instructed and uninstructed ambiguity resolution in relation to depressive symptoms. *Biological Psychology, 109*, 67–72. https://doi.org/10.1016/j.biopsycho.2015.04.007

Savulich, G., Freeman, D., Shergill, S., & Yiend, J. (2015). Interpretation biases in paranoia. *Behavior Therapy, 46*(1), 110–124. https://doi.org/10.1016/j.beth.2014.08.002

Savulich, G., Shergill, S. S., & Yiend, J. (2017). Interpretation biases in clinical paranoia. *Clinical Psychological Science, 5*(6), 985–1000. https://doi.org/10.1177/2167702617718180

Sfärlea, A., Buhl, C., Loechner, J., Neumüller, J., Asperud Thomsen, L., Starman, K., Salemink, E., Schulte-Körne, G., & Platt, B. (2020). "I am a Total…Loser" - the role of interpretation biases in youth depression. *Journal of Abnormal Child Psychology, 48*(10), 1337–1350. https://doi.org/10.1007/s10802-020-00670-3

Spielberger, C. D., Gorsuch, R., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory*. In Palo Alto, CA: Consulting Psychologists Press.

Van Bockstaele, B., Clarke, P. J. F., Notebaert, L., MacLeod, C., & Salemink, E. (2020). Effects of cognitive load during interpretation bias modification on interpretation bias and stress reactivity. *Journal of Behavior Therapy and Experimental Psychiatry, 68*(February), Article 101561. https://doi.org/10.1016/j.jbtep.2020.101561

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viviani, R., Dommes, L., Bosch, J. E., Stingl, J. C., & Beschoner, P. (2018). A computerized version of the scrambled sentences test. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.02310

Wenzlaff, R. M., & Bates, D. E. (1998). Unmasking a cognitive vulnerability to depression: How lapses in mental control reveal depressive thinking. *Journal of Personality and Social Psychology, 75*(6), 1559–1571. https://doi.org/10.1037/0022-3514.75.6.1559

Woud, M. L., Zhang, X. C., Becker, E. S., McNally, R. J., & Margraf, J. (2014). Don't panic: Interpretation bias is predictive of new onsets of panic disorder. *Journal of Anxiety Disorders, 28*(1), 83–87. https://doi.org/10.1016/j.janxdis.2013.11.008

Yiend, J., Allen, P., Lopez, N. D., Falkenberg, I., Tseng, H. H., & McGuire, P. (2019). Negative interpretation biases precede the onset of psychosis. *Behavior Therapy, 50*(4), 718–731. https://doi.org/10.1016/j.beth.2018.11.002

Zahler, L., Meyers, M., Woud, M. L., Blackwell, S. E., Margraf, J., Velten, J., Woud, M. L., Margraf, J., & Velten, J. (2021). Using three indirect measures to assess the role of sexuality-related associations and interpretations for Women's sexual desire: An internet-based experimental study. *Archives of Sexual Behavior, 50*(6), 2471–2484. https://doi.org/10.1007/s10508-020-01897-3

Zahler, L., Sommer, K., Reinecke, A., Wilhelm, F. H., Margraf, J., & Woud, M. L. (2020). Cognitive vulnerability in the context of panic: Assessment of panic-related associations and interpretations in individuals with varying levels of anxiety sensitivity. *Cognitive Therapy and Research, 44*(4), 858–873. https://doi.org/10.1007/s10608-020-10103-6