DOCTORATE IN CLINICAL PSYCHOLOGY

Examining the role of cognitive processes in eating disorders and transdiagnostic repetitive negative thinking

Libby Tindle

Supervised by:

Dr Charlotte Krahé

Dr Frances Meeten

January 2022

Submitted in partial fulfilment of the Doctorate in Clinical Psychology,

University of Liverpool

Acknowledgements

There have been a number of people who have made it possible for me to reach this point. Firstly, I am incredibly humbled and grateful to have been supervised by Dr Charlotte Krahé and Dr Frances Meeten. I have been overwhelmed with how you have both consistently gone above and beyond in your commitment to guiding and supporting me throughout this research. No one could have predicted the challenges that were to come, yet every meeting we had brought encouragement, solutions, and reassurance. Thank you for sharing your knowledge and expertise. I have learnt more than I could have imagined and hope to apply and develop these skills throughout my career.

I would also like to thank Alison Bryant from the Liverpool Experts by Experience team. The enthusiasm you shared and your input with the empirical chapter were invaluable. Thank you for so readily adapting to the changes in the project that none of us could have foreseen. A special thank you as well to my personal tutor, Dr Andrea Flood. Thank you for your kindness, openness, and availability throughout my training experience. You particularly inspired me to keep going, alongside so many professionals and fellow trainees from the Liverpool programme.

Lastly, a personal thank you to my husband, family, and friends. Thank you for your patience and endless support; and believing in me at times when I doubted myself.

Table of Contents

List of Tables

Chapter 1

Chapter 2

List of Figures

Chapter 1

Chapter 2

Word count: 24,834

(exc. Abstracts and references)

Introductory chapter

Thesis overview

This thesis presents two papers that examine the role of cognitive mechanisms involved in the development, maintenance, and treatment of psychopathology. The systematic review (chapter one) provides an overview of reviews that investigate an emerging psychological intervention designed to improve neurocognitive abilities, known as cognitive remediation therapy (CRT; Penadés et al., 2006). CRT was originally developed for use with individuals diagnosed with a brain injury or psychosis (Cicerone et al., 2011; Wykes et al., 2011). The current meta-review focusses on the effectiveness of CRT within eating disorders (EDs). The empirical paper (chapter two) then follows, which explores the role of information processing biases towards understanding the development and maintenance of emotional psychopathology. Specifically, it has sought to build on existing cognitive bias methodologies by psychometrically evaluating an adapted measure of interpretation bias (IB) for worry.

The conceptualisation of mental health difficulties over the last 100 years (and certainly in the West) has predominantly focused on the categorisation within formal taxonomic systems (Kendler, 2009). These taxonomic systems are organised according to hypothetical distinctions between different clusters of signs and symptoms, compiled into a comprehensive collection of psychiatric diagnoses (Dalgleish et al., 2020). The influence of the diagnostic approach to mental health difficulties is widespread and longstanding. This influence includes the application of diagnoses within academic research and clinical practice, the organisation of services within the NHS, as well as being strongly embedded within cultural language. Yet, there is an emerging and growing consensus that psychiatric nosologies may no longer be fit for purpose, both in research and clinical utility (Dalgleish et al., 2020). This consensus has led to the gathering of support for a "transdiagnostic" approach into how we might understand and treat mental health difficulties, which cuts across existing categorical diagnoses, or sets them

aside altogether or sets them aside altogether and focuses on mechanisms that underpin psychological distress (Fusar-Poli et al., 2019). It is proposed that the transdiagnostic approach extends beyond issues of taxonomy and allows for novel ways of thinking about the onset, maintenance, and treatment of psychopathology (Dalgleish et al., 2020). That is, the approach removes the distinctions between psychiatric diagnoses at the level of classification, thus allowing for new ways of conceptualising the underlying processes and theories implicated in mental health difficulties (Dalgleish et al., 2020). This thesis therefore seeks to examine the role of cognitive mechanisms in eating disorders and transdiagnostic repetitive negative thinking.

There has been an increased focus on the role of neurocognitive functioning in EDs in recent years (Stott et al., 2021). Poor treatment outcomes in anorexia nervosa (AN) have led to the development of interventions that focus on addressing both risk and aetiological factors, as well as maintenance models of the condition (Cooper, 2005; Treasure, 2007). Neurocognitive markers of AN include set-shifting difficulties (i.e., difficulties switching between tasks or task demands; Tchanturia et al., 2012) and poor central coherence (i.e., an extreme attention to detail; Lang et al., 2014). Emerging treatments for EDs targeting neurocognitive and executive functions show initial promise (Eichen et al., 2017). These cognitive functions include those sometimes referred to as voluntary "top-down" processes (i.e., task-related effort; attentional control of intrusive thoughts; Hirsch & Mathews, 2012), as targeted in CRT. The majority of research on neurocognitive treatments for EDs focuses on CRT for AN, which has not yet been meta-reviewed. The current review (chapter one) has therefore sought to address this gap in the literature.

In considering the conceptualisation of EDs, Fairburn et al. (2009) developed a transdiagnostic model, contrasting existing disorder-specific models (e.g., AN; bulimia nervosa). The model outlines a set of core processes (i.e., interpersonal difficulties; mood

intolerance; perfectionism; low self-esteem), which are hypothesised in varying combinations to contribute to the maintenance of all EDs. It is argued that the specific features and core processes present in the individual will determine the treatment focus, rather than the diagnosis (Fairburn et al., 2009). This presents a broader shift towards the identification of processes that are common to more than one disorder (Cooper, 2012). Indeed, cognitive theories have proposed that biases in information processing are a key maintaining factor in emotional psychopathology, such as generalised anxiety disorder (GAD; Eysenck, 2013). Worry and anxiety have been recognised in the development and maintenance of EDs, with co-morbidity extremely common (Sternheim et al., 2012; Swinbourne & Touyz, 2007).

Worry is a common experience, and everybody worries from time to time (Ruscio, 2002). It is understood as a form of repetitive negative thinking, characterised by persistent, self-focused, negative, and relatively abstract repetitive thoughts (Krahé et al., 2019; Palmieri et al., 2021). Excessive and uncontrollable worry about a number of future events or activities is a core feature of GAD (American Psychiatric Association, 2013). GAD is associated with a number of difficulties (e.g., stress; avoidance of processing negative emotions) and can have a negative impact on a person's social and adaptive functioning. In understanding worry as a key cognitive symptom of GAD, models of worry have sought to examine the cognitive mechanisms involved in its onset and maintenance (e.g., Hirsch and Mathews, 2012; Davey and Meeten, 2016). These mechanisms include cognitive processing biases, and in particular the tendency to draw negative conclusions in the face of ambiguous information (i.e., IB). As described by Hirsch and Mathews (2012), negative thoughts initially intrude into awareness as the result of an interaction between involuntary "bottom-up" processes (e.g., habitual IB favouring threat content); and voluntary "top-down" processes (e.g., attentional control). There are a range of methods to assess IB; however, a key limitation in this area is the paucity of research examining the reliability and validity of these methods. The current study (chapter

two) has therefore sought to build on existing IB methodologies by psychometrically evaluating an adapted measure of IB, specifically in relation to worry.

Together, these two chapters investigate the role of cognitive mechanisms in both the conceptualisation and treatment of psychopathology. Evidence is growing to suggest that training an individuals' predominant cognitive processes is a promising target in the treatment of emotional and/or eating related difficulties. The current studies contribute to the rigour of psychological research, which is vital towards drawing reliable conclusions and supporting subsequent recommendations within healthcare.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). American Psychiatric Association.

Cicerone, K. D., Langenbahn, D. M., Braden, C., Malec, J. F., Kalmar, K., Fraas, M., Felicetti, T., Laatsch, L., Harley, J. P., & Bergquist, T. (2011). Evidence-based cognitive rehabilitation: updated review of the literature from 2003 through 2008. *Archives of Physical Medicine and Rehabilitation, 92*(4), 519-530.

Cooper, M. (2012). Cognitive behavioural models in eating disorders. In J. Fox & K. Goss (Eds.), *Eating and its Disorders*. John Wiley & Sons.

Cooper, M. J. (2005). Cognitive theory in anorexia nervosa and bulimia nervosa: Progress, development and future directions. *Clinical Psychology Review, 25*(4), 511-531.

Dalgleish, T., Black, M., Johnston, D., & Bevan, A. (2020). Transdiagnostic approaches to mental health problems: Current status and future directions. *Journal of Consulting and Clinical Psychology*, *88*(3), 179.

Davey, G. C., & Meeten, F. (2016). The perseverative worry bout: A review of cognitive, affective and motivational factors that contribute to worry perseveration. *Biological Psychology, 121*, 233-243.

Eichen, D. M., Matheson, B. E., Appleton-Knapp, S. L., & Boutelle, K. N. (2017). Neurocognitive treatments for eating disorders and obesity. *Current Psychiatry Reports, 19*(9), 1-10.

Eysenck, M. W. (2013). *Anxiety: The cognitive perspective*. Psychology Press.

Fairburn, C. G., Cooper D Phil, D. P., Zafra, Doll D Phil, H. A., O'Connor, M. E., Bohn D Phil, D. P., Kristin, Hawker, D. M., Wales, J. A., & Palmer, R. L. (2009).

Transdiagnostic cognitive-behavioral therapy for patients with eating disorders: a two-site trial with 60-week follow-up. *American Journal of Psychiatry, 166*(3), 311-319.

Fusar-Poli, P., Solmi, M., Brondino, N., Davies, C., Chae, C., Politi, P., Borgwardt, S., Lawrie, S. M., Parnas, J., & McGuire, P. (2019). Transdiagnostic psychiatry: a systematic review. *World Psychiatry*, *18*(2), 192-207.

Hirsch, C. R., & Mathews, A. (2012). A cognitive model of pathological worry. *Behaviour Research and Therapy, 50*(10), 636-646.

Kendler, K. S. (2009). An historical framework for psychiatric nosology. *Psychological Medicine*, *39*(12), 1935-1941.

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.

Palmieri, S., Mansueto, G., Scaini, S., Caselli, G., Sapuppo, W., Spada, M. M., Sassaroli, S., & Ruggiero, G. M. (2021). Repetitive negative thinking and eating disorders: A meta-analysis of the role of worry and rumination. *Journal of Clinical Medicine, 10*(11), 2448.

Penadés, R., Catalán, R., Salamero, M., Boget, T., Puig, O., Guarch, J., & Gastó, C. (2006). Cognitive remediation therapy for outpatients with chronic schizophrenia: a controlled and randomized study. *Schizophrenia Research, 87*(1-3), 323-331.

Ruscio, A. M. (2002). Delimiting the boundaries of generalized anxiety disorder: Differentiating high worriers with and without GAD. *Journal of Anxiety Disorders, 16*(4), 377-400.

Sternheim, L., Startup, H., Saeidi, S., Morgan, J., Hugo, P., Russell, A., & Schmidt, U. (2012). Understanding catastrophic worry in eating disorders: process and content characteristics. *Journal of Behavior Therapy and Experimental Psychiatry, 43*(4), 1095-1103.

Stott, N., Fox, J. R., & Williams, M. O. (2021). Attentional bias in eating disorders: A meta-review. *International Journal of Eating Disorders, 54*, 1377-1399.

Swinbourne, J. M., & Touyz, S. W. (2007). The co-morbidity of eating disorders and anxiety disorders: A review. *European Eating Disorders Review: The Professional Journal of the Eating Disorders Association, 15*(4), 253-274.

Treasure, J. L. (2007). Getting beneath the phenotype of anorexia nervosa: the search for viable endophenotypes and genotypes. *The Canadian Journal of Psychiatry, 52*(4), 212-219.

Wykes, T., Huddy, V., Cellard, C., McGurk, S. R., & Czobor, P. (2011). A meta-analysis of cognitive remediation for schizophrenia: methodology and effect sizes. *American Journal of Psychiatry, 168*(5), 472-485.

Chapter One: Systematic review

Cognitive processing in eating disorders:

a meta-review of cognitive remediation therapy and anorexia nervosa

Word count: 9,160

(exc. Abstract and references)

Article intended for submission to the International Journal of Eating Disorders for peer review

Abstract

Objective: There is a clear need for the identification of novel treatments for anorexia nervosa (AN), as reflected in research documenting poor prognosis of the condition. Cognitive remediation therapy (CRT) is a newer psychological intervention that addresses the underlying cognitive mechanisms associated with AN's aetiology and maintenance. The increase in the number of CRT for AN studies has led to a variety of systematic reviews; although the usefulness of the intervention remains in question. This meta-review summarises and synthesises the most reliable findings regarding CRT as a treatment option for AN. **Method:** The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed. Five databases were systematically searched from inception until 15th December 2021, including: Web of Science; SCOPUS; PsycINFO; EMBASE; and PubMed. **Results:** A total of ten systematic reviews (two of which were also meta-analyses) met the eligibility criteria and were retained for inclusion. The quality of each review was assessed using an updated version of the Assessment of Multiple Systematic Reviews (AMSTAR-2; Shea et al., 2017). The identified reviews broadly suggest that CRT is helpful within the treatment of AN; however, the overall classification of the methodological quality for the reviews was critically low, and thus any conclusions drawn should be interpreted with caution. Review characteristics and key findings are summarised. **Discussion:** The most promising findings of CRT for AN relate to improvements in central coherence, cognitive set-shifting, weight-increase reports, and motivation to change; however, effects are modest and further research is needed. Future CRT studies and associated systematic reviews would benefit from a focus on methodological improvement.

Keywords:

Cognitive processing; Anorexia nervosa; Cognitive remediation therapy; Systematic review

1. Introduction

Eating disorders (EDs) are debilitating, intractable conditions that can have a substantial impact on a person's overall health and wellbeing. This includes the diagnosis of anorexia nervosa (AN; American Psychiatric Association, 2013), which is associated with the highest rates of co-morbidity and mortality (Sullivan, 1995; Ulfvebrand et al., 2015). Difficulties with cognitive and emotional functioning are common, with adverse health outcomes indicated even for those only partially meeting diagnostic criteria (Smink et al., 2012; Zipfel et al., 2015). Yet, specific psychological treatments for AN have typically only shown modest improvements (e.g., in terms of clinical course and quality of life; Solmi et al., 2021). There is also a lack of evidence to support superiority or inferiority of the specific treatment options that are available and recommended by clinical guidelines (NICE, 2017; Solmi et al., 2021). As such, there is a clear need for the identification of novel treatments for AN (Berg & Wonderlich, 2013).

A primary goal of research within psychopathology is an understanding of key maintaining processes that can be targeted through interventions. This review focuses on one such intervention, known as cognitive remediation therapy (CRT). In response to the specific cognitive difficulties associated with AN, recent years have seen an increase in research activity for CRT as an intervention for the condition. This has resulted in a number of published reviews; however, there is currently no synthesis of this literature and thus the present paper has sought to provide such an overview.

1.1 Anorexia nervosa conceptualisation and current treatment

AN is characterised by an internalisation of the thin ideal and extreme weight-control behaviours (Hay, 2020). A systematic review by Galmiche et al. (2019) reported weighted population means (and ranges) of lifetime prevalence of AN as 1.4% (0.1–3.6%) for women

and 0.2% (0–0.3%) for men. Indeed, there are reports of increases in AN prevalence in young women, yet the true community incidence of AN remains unknown (Hay, 2020; Smink et al., 2012; Udo & Grilo, 2018). The DSM-V criteria includes the restriction of energy relative to requirements, leading to a significantly low body weight of what would be expected for a person's age, sex, developmental trajectory, and physical health (American Psychiatric Association, 2013). People who do and do not binge-eat or purge (i.e., induce vomiting and/or diuretic misuse) meet criteria, and an atypical AN subtype includes individuals who are not underweight. The diagnosis also includes an intense fear of gaining weight, alongside the person holding a distorted view of themselves (e.g., an overevaluation of body weight or shape), or a denial of the seriousness of the condition (American Psychiatric Association, 2013).

Quality of life has been found to be poor for people meeting diagnostic criteria for AN, with substantial cost placed on individuals, families, and society (Stuhldreher et al., 2012; Zipfel et al., 2015). Although there have been significant advances and developments in the evidence supporting psychological treatment, clinical improvements in AN are modest and relapse rates high (Khalsa et al., 2017). There are a range of NICE (2017) recommended psychological treatments for AN, such as ED-focused cognitive behavioural therapy (CBT-ED) and Maudsley AN Treatment for Adults (MANTRA); the latter being an integrative psychological therapy including principles of motivational interviewing (MI) and cognitive analytic therapy (CAT). For children, treatment recommendations include AN-focused family therapy for children and young people (NICE, 2017). However, there are few treatments for AN that have an established base of empirical support, and high dropout-rates are found in those that have (Berg & Wonderlich, 2013). Treatment effects are often found to be only minimally better than those associated with comparison interventions, such as treatment as usual (TAU) or family therapy (Watson & Bulik, 2013). As such, confusion exists amongst

professionals, patients, families, and providers, with regards to which treatments are useful (Attia, 2021). The present review therefore hoped to alleviate some of the uncertainties relating to CRT, as one of the newer treatments targeting specific cognitive processes. These uncertainties include the potential effectiveness of CRT in terms of setting (e.g., inpatient/outpatient), clinical outcomes (e.g., neuropsychological; ED-related outcomes), mode of delivery (e.g., individual/group; stand-alone or adjunctive), and demographics (e.g., age; diagnostic status). This meta-review should therefore provide a synthesis of existing reviews with varying aims and allow for a comparison between the findings. It would also be helpful to assess the quality of existing reviews, particularly when the evidence may be used as justification for CRT to be offered as a potential treatment option.

1.2 Cognitive processes and anorexia nervosa

There has been a call towards the development of interventions that focus on addressing both risk and aetiological factors, as well as maintenance models of AN (Cooper, 2005; Treasure, 2007). Cognitive theories propose that some traits apparent in the acute phase of AN are predisposing factors (i.e., may be present in childhood), as well as perpetuating factors, thus influencing the course of the condition (Lena et al., 2004). This is supported by research that has found cognitive inefficiencies within the active state of AN, which do not improve following weight-recovery (Green et al., 1996; Szmukler et al., 1992).

There is increasing evidence to suggest executive function deficits across eating and weight disorders (Eichen et al., 2017). Neurocognitive markers of AN include set-shifting difficulties, which involve difficulties switching between tasks or task demands (Tchanturia et al., 2012). Poor central coherence has also been reported (i.e., an extreme attention to detail; Lang et al., 2014). These information processing characteristics, which include cognitive 'rigidity' and perfectionism, are highlighted as possible contributors to the development and

maintenance of the condition (Lopez et al., 2012; Roberts et al., 2013). Individuals with AN also experience impaired emotion recognition, regulation, and expressivity, poor theory of mind, attentional biases, as well as difficulties with socio-emotional processing (Caglar-Nazali et al., 2014; Stott et al., 2021; Zipfel et al., 2015). It is suggested that some neuro- and social cognitive vulnerabilities are present from an early age, which can hinder progress in traditional psychological therapies (Kothari et al., 2013; Southgate et al., 2005).

Emerging treatments for eating and weight disorders targeting neurocognitive and executive functions show initial promise (Eichen et al., 2017). These include those involving involuntary (bottom-up) information processes (i.e., pre-existing processing biases; Hirsch & Mathews, 2012). For example, cognitive bias modification for interpretation biases (CBM-I) target biases towards negative social stimuli (Rowlands et al., 2020; Turton et al., 2018). As well, voluntary (top-down) processes (i.e., task-related effort; attentional control of intrusive thoughts; Hirsch & Mathews, 2012) that translate to clinical interventions may help to reduce perfectionism and cognitive rigidity (Lopez et al., 2012). The latter has been developed by Tchanturia et al. (2007) for the treatment of AN, into the intervention known as CRT. To date, there remains little research evidence within the field of CBM for EDs (see Matheson et al. (2019) for a review on CBM for appearance and self-worth biases in ED psychopathology). The majority of research on neurocognitive treatments for EDs focuses on CRT for AN, which has not yet been meta-reviewed.

1.3 Cognitive remediation therapy for anorexia nervosa

CRT is an umbrella term for psychological interventions that utilise a range of cognitive exercises in order to encourage reflection on an individual's predominant cognitive processes (Lopez et al., 2012). It seeks to facilitate the development of new strategies and thinking skills (e.g., 'thinking about thinking') to enable patients to make behavioural changes within their

everyday life (Tchanturia et al., 2010). The intervention is provided in a variety of ways, depending on the therapist, patient, treatment goals, as well as the programme format (Kim et al., 2018). In general, exercises can be categorised as: (1) those designed to practice and reinforce flexible thinking (set-shifting); and (2) those designed to enhance global processing strategies (coherence; Lopez et al., 2012). There is an emphasis on logic and process, with the exercises also involving other types of basic cognitive functioning (e.g., attention). The method by which CRT may be incorporated into an effective treatment package remains uncertain, both in an inpatient and outpatient setting (Lopez et al., 2012). For example, CRT as a stand-alone treatment might not be sufficient for people with AN who might be very unwell, due to the clinical characteristics of the condition (e.g., poor physical health). Alternatively, a case series by Pitt et al., (2010) suggests that CRT in an outpatient setting should be offered as a pre-therapy intervention. As such, it is likely that CRT is not a stand-alone treatment, but rather a pre-therapy module, or an adjunct to other psychological therapies (e.g., CBT).

CRT is not an intervention unique to AN and was originally developed for use in people diagnosed with a brain injury (Cicerone et al., 2011) and then psychosis (Wykes et al., 2011). Evidence is growing for its application within ED services, specifically in the treatment of AN. This coincides with the growing clinical and empirical interest in understanding neurocognitive functioning in EDs (Smith et al., 2018; Stott et al., 2021). Research suggests that CRT is associated with greater cognitive flexibility for people with a diagnosis of AN, alongside relieving some aspects of eating-related psychopathology and enhancing retention in other therapies (Waller, 2016). The increase in the number of CRT for AN studies has led to a variety of reviews, with the purpose of summarising and synthesizes different researchers' findings (Mulrow, 1994). However, the usefulness of CRT remains in question, and an overview of the best-supported findings across the evidence-base is crucial towards understanding its

suitability and applicability within clinical practice. This could guide and inform further research, as well as current treatment options available within ED services.

1.4 Review aims

There is currently no synthesis of the literature relating to interventions targeting cognitive processes for people with an ED diagnosis and/or related psychopathology. In response to the growing number of systematic reviews in recent years, meta-reviews are a relatively new method of providing such an overview (Stott et al., 2021). The aim of this review of reviews was to (a) broadly characterise and synthesise the relevant literature to date on CRT as an intervention for AN by conducting a systematic review of reviews; (b) describe the methodological standard of the research evidence; and (c) suggest avenues for future research.

2. Method

2.1 Search strategy

The methodology for this review was informed by Smith et al. (2011)'s recommended guidelines for conducting a systematic review of systematic reviews of healthcare interventions. Two search strategies were used. Firstly, five databases were searched from inception until 15th December 2021, including: Web of Science; SCOPUS; PsycINFO; EMBASE; and PubMed. Initial scoping searches were performed to identify the most suitable search terms (see Hagan et al., 2020; Stott et al., 2021; Tchanturia et al., 2017). The search terms used for the review can be found in table 1.

Table 1. Terms used for the search strategy within online databases.

Variable	Search terms used
Review type	systematic OR review OR "meta-analys*"
	AND
Eating disorder	anorex* OR "eating disorder*" OR "disordered eating"
	AND
Intervention	"cognitive remediation therapy" OR cognit* AND (remed* OR train*)

Note. The asterisk symbol allows for identification of terms with the same stem but with ending variation. The Boolean operator AND was used to allow all three variables to be present (and thus focus the search). The Boolean operator OR was used to widen the search.

To ensure that inclusion and exclusion criteria were followed, a second reviewer performed the initial search strategy and reviewed 25% of the titles and abstracts of the articles yielded, following removal of duplicates (N = 61). Full article review was undertaken for 22 of these records. There was agreement on the inclusion and exclusion of 60 out of the 61 articles. The one article where there was disagreement was omitted for inclusion following discussion, whereby it was agreed to not meet eligibility criteria (i.e., a lack of search strategy was provided; see section 2.2).

The second search strategy included hand-searching the reference lists of all included articles, as well as performing the 'cited by' feature on Google Scholar. This was in order to identify any further articles that had not emerged from the database searches.

2.2 Eligibility criteria

To be included, the reviews required the following: (1) available in English language until 13th December 2021, as resources for translation were unavailable; (2) presented in a peer-reviewed and published article or report; (3) article was identified as a systematic review or meta-analysis, including evidence of a search strategy having been performed (ensuring authors tried to capture all relevant information to address their research interests); (4) a description of the search strategy was provided in some detail, even if not full PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses); (5) population includes an existing and/or previous diagnosis of AN (although studies were not required to draw samples from clinical settings); (6) explores three or more intervention studies using CRT across any age group(s) (i.e., child; adolescent; and/or adult). Articles that reviewed fewer than three papers pertaining to CRT were excluded, as this would not have allowed for a synthesis within the existing review. Reviews were also excluded when the focus was on the measurement and/or exploration of cognitive processes and AN. Lastly, articles were excluded when they focused on alternative neurocognitive interventions (e.g., neuromodulation; CBM).

2.3 Quality assessment

The methodological quality of each review was assessed using an updated version of the Assessment of Multiple Systematic Reviews (AMSTAR; Shea et al. 2009). The AMSTAR is a 16-item validated instrument for evaluating the quality of reviews of healthcare interventions and is one of the most widely used tools. It has been found to have superior inter-

rater reliability in relation to another instrument used for the quality assessment of reviews (Pieper et al., 2018). The updated version of the tool (AMSTAR-2) enables more detailed assessment of systematic reviews, including randomised and/or non-randomised studies (Shea et al., 2017). Modifications include simplified response categories, better alignment with the PICOS framework for research questions (i.e., participants, interventions, comparators, outcomes, and study design) and a more detailed consideration of risk of bias (RoB) with included studies. Such adaptations were instigated following relevant critiques of the original AMSTAR (e.g., Faggion, 2015 and Burga et al., 2016). The AMSTAR-2 is also recommended in the Cochrane handbook for conducting meta-reviews for assessing the quality of systematic reviews (Pollock et al., 2018).

The responses to the AMSTAR-2 provide a rating based on the identification of critical domains, or some variation based on these principles, as found in table 2 (Shea et al., 2017). The authors stress that the AMSTAR-2 should not be used to produce an overall score, as this may disguise critical weaknesses that should reduce confidence in the findings (Jüni et al., 1999; Shea et al., 2017). Instead, raters are required to define critical domains of the AMSTAR-2 items prior to commencing a systematic review appraisal (Shea et al., 2017).

Table 2. Rating overall confidence in the results of the review using the AMSTAR-2 tool (Shea et al., 2017).

Review Rating	Description
High	No or one non-critical weakness: the systematic review provides an accurate and comprehensive summary of the results of the available studies that address the question of interest.
Moderate	More than one non-critical weakness †: the systematic review has more than one weakness but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review.
Low	One critical flaw with or without non-critical weaknesses: the review has a critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest.
Critically low	More than one critical flaw with or without non-critical weaknesses: the review has more than one critical flaw and should not be relied on to provide an accurate and comprehensive summary of the available studies.

† *Multiple non-critical weaknesses may diminish confidence in the review, and it may be appropriate to move the overall appraisal down from moderate to low confidence.*

The identification of critical items differs between AMSTAR-2 users, according to the specific research question (Shea et al., 2017). The present study identified seven domains that might seriously affect the validity of the findings and conclusions of the included review should they not be demonstrated (i.e., a 'critical weakness'). These critical domains were used to appraise review quality and are outlined in the table 3 (including both critical/non-critical items). Whilst non-critical items also indicate review quality, they are considered less harmful to its validity should they not be present (i.e., a 'non-critical weakness').

Table 3. AMSTAR-2 items and the identification of critical/non-critical domains.

	AMSTAR-2 item	Critical? (Y/N)
(1)	The research questions and inclusion criteria for the review included the components of PICO	N
(2)	Protocol registered before commencement of the review †	N
(3)	The authors explained their selection of the study designs for inclusion in the review	N
(4)	Adequacy of the literature search †	Y

(5)	The authors performed study selection in duplicate	N
(6)	The authors performed data extraction in duplicate	N
(7)	Justification for excluding individual studies †	Y
(8)	Description of included studies in adequate detail	N
(9)	Satisfactory technique for assessing risk of bias from individual studies being included in the review †	Y
(10)	The authors reported on the sources of funding for included studies	N
(11)	Appropriateness of meta-analytical methods †	Y
(12)	If meta-analysis was performed, the review authors assessed the potential impact of risk of bias in individual studies on the results of the meta-analysis or other evidence synthesis	N
(13)	Consideration of risk of bias in individual studies when interpreting the results of the review †	Y
(14)	Satisfactory explanation for any heterogeneity observed in the review's results	N
(15)	Assessment of presence and likely impact of publication bias †	Y
(16)	The review authors reported any potential sources of N conflict of interest, including any funding to conduct the review	N
(17)	The review authors considered the impact that the Y sample sizes of included studies may have had on the reliability of the review's findings	Y

† *Recommended critical domains according to the developers of the tool.*

The current study agreed with the applicability of the recommended critical items by Shea et al. (2017) for the present meta-review, as indicated in table 3. For example, items 9 and 13 consider whether RoB in primary studies was assessed, which determines the degree to which the results can be relied upon (Stott et al., 2021). Consideration of the possibility of publication bias would also affect validity (item 15), as well as whether meta-analysis (if performed) used appropriate methods for statistical combination of results (item 11). The latter

would seek to reduce any bias in effect size estimates. It was also agreed that minimising selection bias should be critical (as per items 4 and 7), as this could undermine validity of the conclusions drawn. Similar to a meta-review by Stott et al. (2021) and despite the recommendations from Shea et al. (2017), item 2 (protocol registered) was not included as a critical domain, as this would have affected the assessment of quality variability. Indeed, none of the reviews included in this meta-review met this criterion. The current study also took guidance from Stott et al. (2021) and added a critical item to assess whether the review considered the influence of small studies on the reliability of its findings (item 17).

A copy of the AMSTAR-2 can be found in Appendix A, which includes additional details for completing the items.

2.4 Risk of bias

To assess the inter-rater reliability of the AMSTAR-2 ratings, 30% of all included reviews were quality appraised by a second reviewer (N = 3). Intra-class correlation coefficients (ICC) were computed, and a high degree of inter-rater reliability was found for the items between the raters. The average measure ICC was .83 with a 95% confidence interval from .69 to .90 (F (47, 47) = 5.771, p <.000). A two-way mixed effects model was performed with absolute agreement definition (i.e., column variance is excluded from denominator variance). Any disagreements with regards to review quality appraisal were discussed until a consensus was reached.

2.5 Review selection

The review search process is outlined in the PRISMA flow diagram within the Results section (see figure 1; Moher et al. 2009). The reference management software Endnote was used for direct exportation of citations from the online databases. Each database was organised

into separate reference manager files, alongside any additional records identified. The search results were combined, and duplicates removed. Titles and abstracts were initially screened for relevance against the inclusion and exclusion criteria. Full texts were sourced to establish eligibility when it was not clear from the abstract. The same criteria were used during full-article review. Consultation with the research team was sought for selection of the final articles and any discrepancies resolved via discussion, alongside input from the second reviewer.

2.6 Data extraction and synthesis

A checklist was developed in order to systematically extract and summarise key information and results from each review. This included: (a) authors and publication year; (b) method; (c) location; (d) conflicts of interest; (e) review aims; (f) number of relevant studies (including total number of participants); (g) sample demographics; (h) diagnostic status; (i) type of intervention (e.g., stand-alone, or adjunctive); (j) comparison group; (k) outcome measures used; (l) key findings. After the final articles were identified, relevant data was synthesised according to the guidelines. When information beyond the interest of this review was included within the article, only the relevant information was extracted.

The heterogeneous nature of the extracted data and the broad nature of the study aims meant that statistical techniques were not appropriate for synthesis of the literature. Informal analyses that seek to summarise the results could also introduce significant overlap, as well as primary studies being included more than once, leading to biased results (Pieper et al., 2014). A narrative synthesis was therefore conducted. The AMSTAR-2 ratings were completed following the checklist.

3. Results

3.1 Search results

This meta-review was conducted using the PRISMA statement (Liberati et al., 2009). A flow diagram of the study selection process is shown in figure 1. The initial online database search identified a total of 385 records. Following the removal of duplicates, 243 of these records remained. The titles and abstracts of these were screened for relevance and a further 184 records were excluded for not meeting eligibility criteria. Common reasons for exclusion at this stage were the article being a non-review paper and/or the focus being on other treatment modalities. The records yielded by reference list searching were excluded following full article review, due to there being no reference to CRT and rather a focus on other treatment modalities. A list of the full article reviewed and excluded records can be found in appendix B. A total of ten articles proved to fulfil the eligibility criteria and were retained for inclusion.

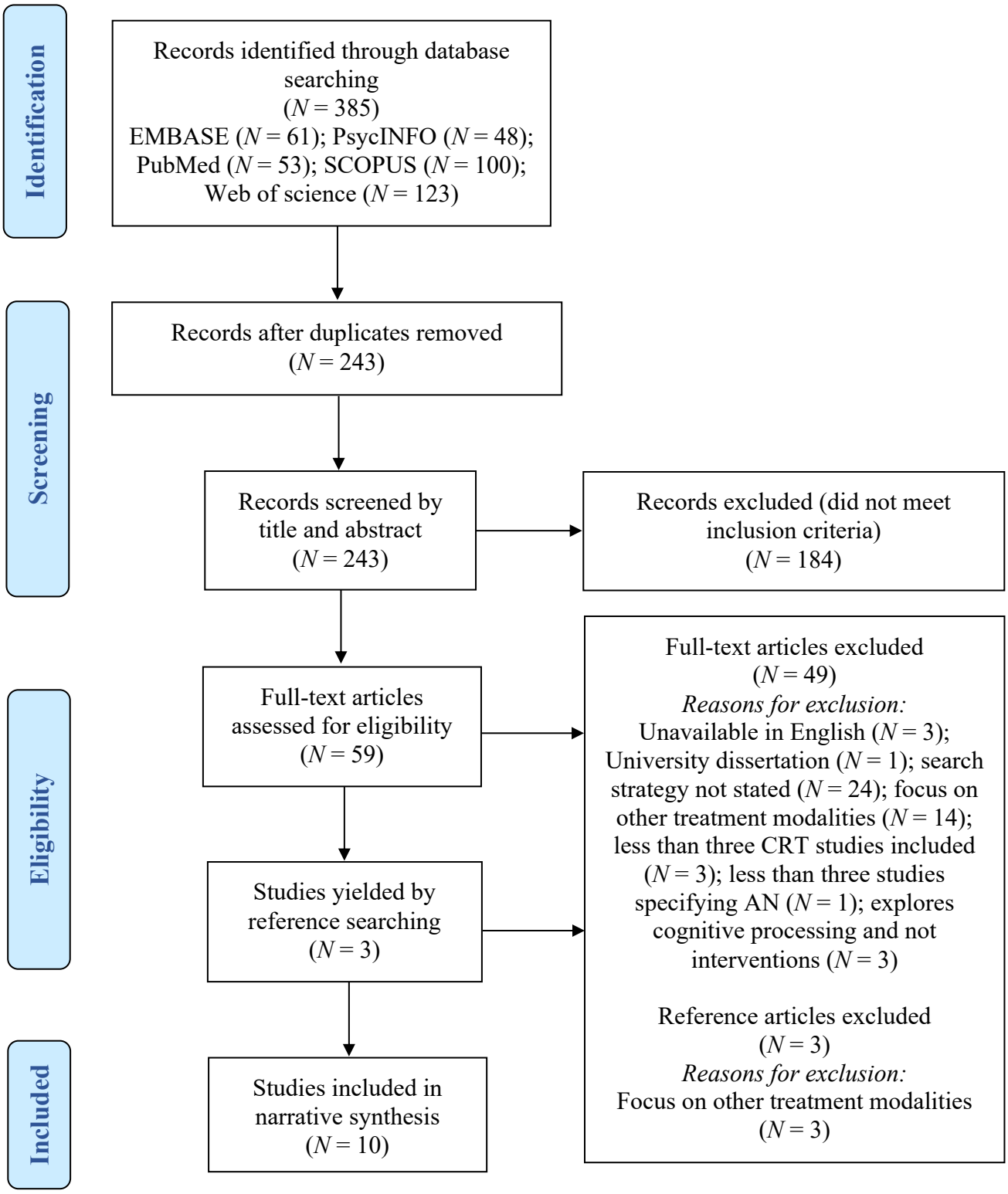


Figure 1. PRISMA (2009) flow diagram of study selection.

3.2 Methodological quality of reviews

Each of the included reviews were assigned one of the four methodological quality ratings, as per the AMSTAR-2 guidance (i.e., “high”, “moderate”, “low”, “critically low”; see section 2.2 for further detail). The ratings can be found in table 4, including a breakdown of each item response.

Table 4 shows the classification of the methodological quality for eight of the reviews as “critically low”, as well as “low” for the remaining two. In terms of critical domains, there were only three studies that did not report an adequate literature search (item 4). Most of the reviews provided adequate detail in terms of excluded studies and justified these exclusions (item 7). Four of the reviews reported the use of a satisfactory technique for assessing RoB (item 9); however, only two of these studies made accounts for RoB when interpreting and/or discussing the results (item 13). The studies that performed a meta-analysis (N = 2) used appropriate methods for statistical combination of the results (item 11) and carried out an adequate investigation of publication bias (item 15).

In considering non-critical items, only one of the reviews failed to describe the research question(s) and inclusion criteria in a way that covered the components of PICO (item 1). None of the included reviews contained a statement to indicate that the methodology was established in priori (item 2). Most studies explained their selection of the study designs for inclusion in the review (item 3) and discussed heterogeneity observed in the results (item 14). Half of the studies reportedly performed study selection and data extraction in duplication (items 5 and 6). There was adequate detail of the studies included in nine of the reviews, although this varied in terms of the amount of detail provided (item 8). None of the reviews reported on the sources of funding for the studies included (item 10); although most of the reviews reported on their own potential conflicts of interest (item 16). The two studies that performed a meta-analysis did not assess the potential impact of RoB in individual studies on the results (item 12).

Table 4. AMSTAR-2 quality assessment for included reviews.

AMSTAR-2 question no. †		Included reviews									
		Brockmeyer et al. (2018)	**Couturier et al. (2020)**	**Dahlgren and Rø (2014)**	**Denison-Day et al. (2018)**	**Gramaglia et al. (2020)**	**Hagan et al. (2020)**	**Isserlin et al. (2020)**	**Sala et al. (2016)**	**Tchanturia et al. (2017)**	**Tchanturia et al. (2014)**
(1)	PICO	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
(2)	Protocol	N	N	N	N	N	N	N	N	N	N
(3)	Inclusion	Y	Y	Y	Y	N	Y	Y	Y	Y	N
(4) ‡	Search strategy	Y-(P)	Y-(P)	Y-(P)	Y	N	Y-(P)	N	N	Y-(P)	Y-(P)
(5)	Selection in duplicate	N	Y	N	Y	Y	Y	Y	N	N	N
(6)	Extraction in duplicate	Y	N	N	Y	Y	Y	Y	N	N	N
(7) ‡	Exclusion justification	N	Y	Y	Y-(P)	Y-(P)	Y-(P)	Y-(P)	Y-(P)	Y-(P)	N
(8)	Study detail	Y-(P)	Y	Y-(P)	Y-(P)	Y-(P)	Y	Y-(P)	Y	Y-(P)	N
(9) ‡	Risk of bias technique	N	Y	N	Y	N	N	Y	N	Y	N
(10)	Funding	N	N	N	N	N	N	N	N	N	N
(11) ‡	Statistical methods	N/A	N/A	N/A	N/A	N/A	Y	N/A	N/A	Y	N/A
(12)	Meta-analysis	N/A	N/A	N/A	N/A	N/A	N	N/A	N/A	N	N/A
(13) ‡	Risk of bias interpretation	N	N	N	Y	N	N	Y	N	N	N

		Brockmeyer et al. (2018)	Couturier et al. (2020)	Dahlgren and Rø (2014)	Denison-Day et al. (2018)	Gramaglia et al. (2020)	Hagan et al. (2020)	Isserlin et al. (2020)	Sala et al. (2016)	Tchanturia et al. (2017)	Tchanturia et al. (2014)
(14)	Heterogeneity observed	N	Y	Y	Y	Y	Y	Y	N	Y	Y
(15) ‡	Publication bias	N/A	N/A	N/A	N/A	N/A	Y	N/A	N/A	Y	N/A
(16)	Conflict of interest	Y	Y	Y	N	Y	Y	Y	Y	N	N
(17) ‡	Sample size	N	N	Y	N	N	Y	N	Y	Y	Y
Quality category §		**Critically low** C = 4 NC = 4	**Critically low** C = 2 NC = 3	**Critically low** C = 2 NC = 4	**Low** C = 1 NC = 3	**Critically low** C = 4 NC = 3	**Critically low** C = 2 NC = 3	**Critically low** C = 2 NC = 2	**Critically low** C = 3 NC = 5	**Low** C = 1 NC = 6	**Critically low** C = 3 NC = 8

Note. No. = number; Y = yes; N = no; P = partial; C = number of critical weaknesses; NC = number of non- critical weaknesses.

† ***(1)*** *Did the research questions and inclusion criteria for the review include the components of PICO?* ***(2)*** *Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?* ***(3)*** *Did the review authors explain their selection of the study designs for inclusion in the review?* ***(4)*** *Did the review authors use a comprehensive literature search strategy?* ***(5)*** *Did the review authors perform study selection in duplicate?* ***(6)*** *Did the review authors perform data extraction in duplicate?* ***(7)*** *Did the review authors provide a list of excluded studies and justify the exclusions?* ***(8)*** *Did the review authors describe the included studies in adequate detail?* ***(9)*** *Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?* ***(10)*** *Did the review authors report on the sources of funding for the studies included in the review?* ***(11)*** *If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?* ***(12)*** *If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?* ***(13)*** *Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?* ***(14)*** *Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?* ***(15)*** *If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?* ***(16)*** *Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?* ***(17)*** *Did the review authors consider the impact that the sample sizes of included studies may have had on the reliability of the review's findings?*

‡ *Identified critical domains.*

§ *AMSTAR-2 quality category (i.e., "high" – review met all critical domains and failed on no more than three non-critical domains; "moderate" – review did not meet more than three non-critical domains but met all critical domains; "low" – review did not meet one critical domain; "critically low" – review did not meet more than one critical domain).*

3.3 Overlap between reviews

The degree of overlap between the relevant publications within all the systematic reviews that met eligibility criteria was calculated using the "corrected covered area" (CCA; Pieper et al., 2014). The CCA determines the frequency of repeated inclusion of studies, offering a percentile representation of the degree of overlap (i.e., the higher the percentage the greater the overlap). It is calculated by dividing the frequency of repeated occurrences of index publications amongst the reviews (subtracted by the number of index publications) by the product of the number of index publications and reviews, which is reduced by the number of index publications (i.e., 86 – 54/54 x 10 – 54; Pieper et al., 2014). The current meta-review found the overlap to be 6.6%, which is deemed to fall within the moderate range (i.e., between 6% and 10%; Pieper et al., 2014). The included reviews thus provide an acceptable degree of independent information for overview. The total number of primary articles reviewed by included reviews (without duplicates) was 54.

3.4 Review characteristics

The following information regarding review characteristics can be found in table 5, including authors and publication year, methodology, aims, location, funding/conflicts of interest, as well as the number of relevant studies included. All included articles were published between 2014 and 2020. Nine of the articles were identified by the authors as a "systematic review". Gramaglia et al. (2020) referred to their paper as a "scoping review", which met criteria due to having conducted a systematic search strategy. Two of the articles also included a meta-analysis. Half of the reviews were conducted in the United Kingdom (UK), with three of these co-produced within institutions in Germany, Norway, and the United States of America (USA). The other half were conducted in the USA (N = 2), Canada (N = 2) and Italy (N = 1). Sources of funding that were reported included University institutions, health services

Table 5. Extracted data – review characteristics.

Included review	Review characteristics						
	Method	**Location**	**Sources of funding**	**Conflict of interest**	**Aim(s) of review**	**No. relevant studies included in meta-review**	**No. participants in relevant studies; total † (range)**
Brockmeyer et al. (2018)	SR	Düsseldorf, Germany; London, UK	Last author is supported by a NIHR Senior Investigator Award and receives salary support from the NIHR Mental Health Biomedical Research Centre at the South London and Maudsley NHS Foundation Trust and KCL	None	(1) To summarise the main findings of clinical trials on established AN treatments since a 2013 review – 'established' treatments are those that are widely used in AN treatment, recommended by guidelines and/or have been tested in at least one large RCT, with a minimal sample size of $N = 100$); (2) to review associated process outcome studies, which shed light on active ingredients of existing treatment approaches; (3) to review emerging treatments for AN, i.e. those that have only been (or are currently being) tested in feasibility or pilot trials so far; (4) to attempt to forecast future developments by assessing forthcoming/unpublished trials.	5 published trials and 7 unpublished trials (out of 30 registered trials and 34 registered/ unpublished trials)	807 (30-168)
Couturier et al. (2020)	SR	Hamilton, Canada	The Canadian Institutes for Health Research	None	To synthesise the best available evidence on treatments for children and adolescents with EDs resulting in the production of a practice guideline.	8 (out of 48 for individual and group psychotherapy PRISMA)	61 (1-125)

	Method	Location	Sources of funding	Conflict of interest	Aim(s) of review	No. relevant studies included in meta-review	No. participants in relevant studies; total † (range)
Dahlgren and Rø (2014)	SR	Oslo, Norway; Oxford, UK.	The Centre for Research on EDs at Oxford University (CREDO); the Regional Department for ED, Department of Mental Health and Addiction (Oslo University Hospital, Ullevål)	None	To present a systematic overview of all studies of CRT for AN, in order delineate the interventions developmental trajectory in both adults and adolescents, its current state, and its potency in future treatment of AN.	21 (out of 21)	362 (1-46)
Denison-Day et al. (2018)	SR	Bournemouth, UK	Bournemouth University	Not reported	To conduct a systematic review to see how effective specific motivation-focused therapies are at improving motivation to change in relation to other therapies. To answer the following: (1) what evidence is there that treatment interventions increase motivation to change amongst individuals with EDs? (2) are specifically motivationally focused interventions more effective than established treatments at increasing motivation to change?	6 (out of 41)	282 (7-98)
Gramaglia et al. (2020)	Scoping review/ systematic search strategy	Novara, Italy	Not reported	None	To assess treatment outcome in AN in terms of changes in alexithymia as assessed by the TAS (most used self-report measure).	3 (out of 11)	131 (32-66)

	Method	Location	Sources of funding	Conflict of interest	Aim(s) of review	No. relevant studies included in meta-review	No. participants in relevant studies; total † (range)
Hagan et al. (2020)	SR and MA	USA	The 2017 University of Kansas Graduate Research Summer Scholarship (awarded to Dr Hagan)	None	(1) To conduct a SR and MA of RCTs of CRT for AN on neuropsychological (i.e., set-shifting, central coherence) outcomes at EoT. (2) to examine the effect of CRT for AN on patient dropout, ED-related, and other psychological outcomes at EoT.	9 (out of 9)	611 (30-275)
Isserlin et al. (2020)	SR	Canada	No funding sources	None	To explore the scope and benefits of psychological treatments provided to children and adolescents with EDs in inpatient settings at the time of hospital discharge.	4 (out of 66)	127 (2-70)
Sala et al. (2016)	SR	USA	Not reported	None	To identify, summarize, and evaluate emotion-focused treatments for AN. Specifically, to provide an overview of various emotion-focused treatments and assesses their acceptability, feasibility, and efficacy.	5 (out of 20)	218 (1-81)
Tchanturia et al. (2017)	SR and MA	UK	Not reported	Not reported	To synthesise the existing evidence for the effects of CRT in YP with AN whilst comparing these findings to research in adult populations.	9 (out of 9)	435 (1-173)
Tchanturia et al. (2014)	SR	UK; Georgia	Maudsley Charity Health in Mind	Not reported	To appraise the current evidence-base across four conditions (AN/MDD/OCD/ASD) and to consider any similarities and differences in the use of such approaches. To focus on how the	15 (out of 24)	Not reported in full (1-82 approx.)

				main characteristics and research findings from different treatment approaches in other conditions could be related to AN.

Note. No. = number; SR = systematic review; meta-analysis; UK = United Kingdom; USA = United States of America; NIHR = National Institute of Health Research; ED = eating disorder; KCL = King's College London; AN = anorexia nervosa; RCT = randomised controlled trial; CRT = cognitive remediation therapy; TAS = Toronto Alexithymia Scale; EoT = end of treatment; YP = young people; MDD = major depressive disorder; OCD = obsessive-compulsive disorder; ASD = autism spectrum disorder.

† *Including control.*

within the UK and Canada, a mental health charity, as well as a graduate research scholarship in the USA. Three of the studies did not report any sources of funding. Seven of the studies reported no conflicts of interests, whilst three studies did not comment on potential conflicts of interest.

The number of relevant studies included within the reviews ranged from 3 to 15, with the total number of participants ranging from 61 to 807 in each review. The types of studies included in each review varied within and between each review, including case studies and randomised control trials (RCTs). As such, there was variation in terms of the number of participants included in the individual studies. Brockmeyer et al. (2018) was the only review that included registered/unpublished studies, as opposed to published studies.

Dahlgren and Rø (2014), Tchanturia et al. (2017) and Hagan et al. (2020) specifically focused on the potential of CRT as an intervention for AN. An earlier review by Tchanturia et al. (2014) appraised the evidence-base of CRT across AN and related conditions (e.g., major depressive disorder), in order to see if the findings could be related to AN. Brockmeyer et al. (2018) reviewed emerging treatments for AN (e.g., CRT; neuromodulation; exposure-based trials) amongst those that are already established (e.g., family interventions; MANTRA). Couturier et al. (2020) and Isserlin et al. (2020) both sought to review the evidence-base on the various treatments for children and adolescents with EDs. Isserlin et al. (2020) specifically explored interventions provided in inpatient settings at the time of hospital discharge. Couturier et al. (2020) produced the first Canadian Practice Guidelines to evaluate the evidence on psychotherapeutic and psychopharmacological treatments. Other reviews focused on the evidence of interventions for specific outcomes, such as motivation to change amongst individuals with EDs (Denison-Day et al., 2018). Gramaglia et al. (2020) assessed treatment outcomes in AN in terms of changes in alexithymia. Lastly, Sala et al. (2016) evaluated emotion-focused treatments for AN, which included cognitive remediation and emotion skills training (CREST). Overall, there was a range of outcomes focused on in each review, with

some reviews specifying the measures of interest, and thus a heterogenous set of studies reported.

3.5 Summary of key findings

The following synthesis of information regarding the findings from relevant studies (e.g., sample demographics, intervention, outcome measures) and review conclusions can be found in table 6. Key qualitative information with regards to conclusions drawn have also been integrated within the synthesis. Table 6 has classified the findings reported in terms of overall helpfulness of CRT (i.e., 'yes' or 'no'). When the majority of conclusions drawn within the review favoured the use of the therapy (even when limitations and/or poor outcomes were reported), then it was deemed to have found CRT helpful (i.e., rated as 'yes'). CRT was identified as not helpful when the review mostly reported findings that did not support its effectiveness (i.e., rated as 'no'). For the purpose of this meta-review, the summary does not report on findings related to non-relevant studies (e.g., those evaluating the efficacy of other types of interventions).

3.5.1 Sample demographics. Three of the reviews included child and adolescent samples in the included studies (Couturier et al., 2020; Isserlin et al., 2020; Tchanturia et al., 2017) and one of the reviews included only adults (Tchanturia et al., 2014). The Gramaglia et al. (2020) review included studies with child, adolescent and adult samples, and the remaining studies included adult and adolescents. As such, there was generally a broad range of age groups reported within the reviews; however, key findings and conclusions typically differed. For example, CRT was identified as a novel and emerging treatment for AN, with Dahlgren and Rø (2014) concluding that there is strong evidence of its feasibility across ages and AN illness severity. On the other hand, Isserlin et al. (2020) explained that due to the quality of the studies identified, it was not possible to draw any conclusions about the effectiveness of CRT for children and young people. This was similar for the other adjunctive treatments examined within the review (e.g., bright light therapy). Indeed, the quality ratings for both Dahlgren and

Table 6. Extracted data – findings from relevant studies and the review conclusions.

Included review	Findings from relevant studies †						Review conclusions
	Sample demographics (sex; ethnicity; age; SES)	**Diagnostic status**	**Intervention (session no. range) and setting/format**	**Stand-alone or adjunctive**	**Comparison group**	**Outcome measures**	**CRT helpful (yes/no) ‡**
Brockmeyer et al. (2018)	Sex: Female 100% (*N* = 3); Ethnicity: NR; Age: adult/ adolescent (*N* = 5), adult (*N* = 3), adolescents (*N* = 3), NR (*N* = 1); SES: NR.	Former AN (weight-restored) (*N* = 1); AN (*N* = 8); AN-R (*N* = 1); BMI < 19 + ED diagnosis (*N* = 1); AN or OCD (*N* = 1).	CRT as an adjunct to other therapies (6-30). Inpatient; outpatient; individual; group.	Adjunct to TAU (*N* = 7); Adjunct to CBT (*N* = 3); Adjunct to FBT (*N* = 1); Adjunct to behaviour therapy (*N* = 1).	TAU (*N* = 3); 24 session CBT (*N* = 1); 30 sessions NNT + TAU (*N* = 1); 10 sessions art therapy + TAU (*N* = 1); 10 sessions SHAM + TAU (*N* = 1); 10 sessions supportive counselling therapy + CBT (*N* = 1); 10 sessions relaxation therapy + TAU (*N* = 1); 6 sessions CBT only (*N* = 1); FBT + art therapy (15 sessions of each) (*N* = 1); 12 sessions CRT + Behaviour Therapy (*N* = 1).	ED-related QoL; ED-related psychopathology; BMI; set-shifting, central coherence + emotion recognition; Drop-out rate; Neuropsychological functioning + clinical outcomes (Morgan– Russell Scales); treatment engagement + motivation; EDE-Q; cognitive flexibility; full remission from AN; calorie intake; attrition rate.	Yes
Couturier et al. (2020)	Sex: Female 90%, male 10% (*N* = 2); Ethnicity: NR; Age: child/adolescents (*N* = 8). SES: NR.	AN	CRT as an adjunct to other therapies (10-15). Inpatient; outpatient; individual; group.	Adjunct to FBT (*N* = 1); Adjunct to multimodal inpatient treatment (*N* = 7).	Art therapy (*N* = 2); NR (*N* = 2); TAU (*N* = 1); No control group (*N* = 3).	Global EDE-Q score; BMI; depression scores; anxiety scores; weight improvement; EBRS scores; EDE-Q; MSCARED; EAT; EAT-26.	No

	Sample demographics (sex; ethnicity; age; SES)	Diagnostic status	Intervention (session no. range) and setting/format	Stand-alone or adjunctive	Comparison group	Outcome measures	CRT helpful (yes/no) ‡
Dahlgren and Rø (2014)	Sex: Female 98%, male 2% (*N* = 1); Ethnicity: NR; Age: adult/adolescent (*N* = 9), adult (*N* = 11); SES: NR.	AN; AN + EDNOS; AN + recovered AN.	CRT (4-30). Inpatient; outpatient; individual; group; family; computer assisted + individual.	Stand-alone (*N* = 21).	NR (*N* = 18); AN-EXRP (*N* = 1); TAU (*N* = 1); NNT (*N* = 1).	Cognitive set-shifting; BMI; cognitive flexibility + positive feedback; perfectionism; Brixton and CatBat tasks (cognitive performance); self-reports of ability to change; neuropsychological performance; weight, depression, visuospatial memory, global information processing + verbal fluency; patient brief shift subscale; short-term attrition; cognitive set-shifting; treatment acceptance; better caloric intake; ED-related QoL and ED psychopathology.	Yes
Denison-Day et al. (2018)	Sex: Female 98.75%, male (1.25% (*N* = 4); Ethnicity: NR; Age: adult (*N* = 4), adolescent (*N* = 2); SES: NR.	AN, BN, EDNOS	CRT; MDT treatment programme + CRT (4-10). Individual; group.	Adjunct to MDT treatment programme (*N* = 1); stand-alone (*N* = 5).	TAU (*N* = 1); No control (*N* = 4); Low and high autism scores (*N* = 1).	Motivational ruler.	Yes
Gramaglia et al. (2020)	Sex: Female 95.9%, male 4.1% (*N* = 2); Ethnicity: NR; Age: child/adolescents (*N* = 3), adult (*N* = 2). SES: NR.	AN	Manualised CREST (5-10). Inpatient; individual; group.	Adjunct to Emotion Skill Training (*N* = 3).	NR (*N* = 3).	RSAS; TAS; Motivational ruler; SQ; ERQ-CA; AQ-10; BMI.	Yes

	Sample demographics (sex; ethnicity; age; SES)	Diagnostic status	Intervention (session no. range) and setting/format	Stand-alone or adjunctive	Comparison group	Outcome measures	CRT helpful (yes/no) ‡
Hagan et al. (2020)	Sex: Female 94%, male 6% (*N* = 9); Ethnicity: NR; Age: adult (*N* = 7), adolescent (*N* = 2). SES: NR.	AN-B; AN or BN; NR	CRT; CRT as an adjunct to other therapies (6-12). Inpatient; outpatient.	CRT only; CRT + CBT; CRT + TAU; CRT + FBT.	CBT (*N* = 1); NNT + TAU (*N* = 1); TAU (*N* = 3); AN-EXRP (*N* = 1); art therapy + FBT (*N* = 1); SAT + TAU (*N* = 1).	RCFT; D-KEFS; CWIT; BMI; EDE-Q; BDI; cued task switching paradigm; Visual target-detection task; RCFT; WCST; ED-related QoL; calories consumed in test meal; BRIEF-SR; ChEDEQ; DIKJ; STAI(C)-T; YBC-EDS; CY- BOCS; CTMT; thought records.	No
Isserlin et al. (2020)	Sex: NR; Ethnicity: NR; Age: child/adolescents (*N* = 4); SES: NR.	AN-R; AN; AN-B; atypical AN	CRT as an adjunct to other therapies (4-10). Inpatient.	Adjunct to multimodal inpatient treatment (*N* = 4).	TAU (*N* = 1); No control (*N* = 3).	BMI; EAT-26; EBRS; %TGW; EDE-Q; motivational stage of change.	No
Sala et al. (2016)	Sex: Female 100% (*N* = 1); Ethnicity: NR; Age: adult (*N* = 4), adolescent/adult (*N* = 1); SES: NR.	AN; AN-R	CREST (5-10. Inpatient.	CREST (*N* = 4); CREST + TAU (*N* = 1)	TAU (*N* = 1); No control (*N* = 4).	BMI; EDE-Q; emotion reappraisal; TAS; set-shifting tasks; RSAS.	No
Tchanturia et al. (2017)	Sex: NR; Ethnicity: NR; Age: child (*N* = 1), adolescent (*N* = 7); SES: NR.	AN	CRT (8-10). Individual; group.	CRT only; CRT + TAU.	TAU (*N* = 1); HC not receiving CRT (*N* = 1); No control (*N* = 6).	ROCFT; TMT-4; BRIEF-SR; feedback letters; CWIT; Tower Test; 'The Ravello Profile'; the GEFT; D-KEFS (letter fluency); Cognitive Flexibility Scale; Motivational Ruler; SQ.	Yes
Tchanturia et al. (2014)	Sex: NR; Ethnicity: NR; Age: adult (*N* = 15); SES: NR.	AN	CRT. Inpatient; outpatient; individual; group.	NR; CRT + TAU; manual-based CRT.	NR (*N* = 10); TAU (*N* = 1); NNT (*N* = 1); AN-EXRP (*N* = 1); CBT (*N* = 1).	Neuropsychological measures of set-shifting; patient feed-back; cognitive flexibility task; central coherence; BMI; depression symptoms; dropout	Yes

										rate; general functioning; QoL; computer-based task-switching paradigm; caloric intake at a test meal; eating-related anxiety; cognitive flexibility, self-esteem; motivation to change.

Note. No. = number; AN = anorexia nervosa; NR = not reported; SES = socioeconomic status; AN-R = anorexia nervosa (restrictive subtype); BMI = body mass index; OCD = obsessive-compulsive disorder; EDNOS = eating disorder not otherwise specified; BN = bulimia nervosa; AN-B = anorexia nervosa-binge eating/purging sub type; CRT = cognitive remediation therapy; MDT = multidisciplinary team; CREST = cognitive remediation and emotion skills training; TAU = treatment as usual; CBT = cognitive behavioural therapy; FBT = family-based treatment; NNT = non-specific neurocognitive therapy; SHAM = physical activity, emotion recognition, interpersonal functioning; AN-EXRP = exposure and response prevention for anorexia nervosa; SAT = aims to help increase positive experiences through various exercises, such as playing board game; HC = healthy control; QoL = quality of life; EBRS = the eating disorder rating scale; EDE-Q = eating disorder examination questionnaire; MSCARED = motivational stages of change for adolescents recovering from an eating disorder; EAT = eating attitudes test; EAT-26 = eating attitudes test-26 item; RSAS = revised social anhedonia scale; TAS = toronto alexithymia scale; SQ = satisfaction questionnaire; ERQ-CA = emotion regulation questionnaire for children and adolescents; AQ-10 = the autism spectrum quotient-10 items; RCFT = the Rey complex figure test*; D-KEFS = the delis–kaplan executive function system; CWIT = colour-word interference test; BDI = beck depression inventory; WCST = the* Wisconsin card sorting test*; BRIEF-SR = behaviour rating inventory of executive function; ChEDEQ = child eating disorder examination questionnaire; DIKJ = depression inventory for children and adolescents; STAI(C)-T = the state-trait anxiety inventory for children; YBC-EDS = the Yale-Brown-Cornell eating disorders scale; CY- BOCS = the children's Yale-Brown obsessive compulsive scale; CTMT = comprehensive trail-making test; TGW = treatment goal weight; ROCFT = the Rey-Osterrieth complex figure test; TMT-4 = trial making test condition 4; GEFT = the* group embedded figures test*.*

† *N refers to no. of studies (provided when information available).*

‡ *When CRT findings were variable (i.e., reports of both significant and non-significant outcomes), then the yes/no response is a majority rating (e.g., yes = more helpful findings reported). This is a crude global recommendation and only meant as an indicative conclusion (e.g., it does not take context into account). See section 4.3 for a discussion of the limitations of this approach.*

Rø (2014) and Isserlin et al. (2020) were critically low and any conclusions drawn should be interpreted with caution. It was thus not possible to differentiate effectiveness of CRT between age groups.

Hagan et al. (2020) was the only study that reported the gender of participants in all the studies, which equated to 94% female and 6% male. Couturier et al. (2020) and Gramaglia et al. (2020) reported gender in two of the studies, both identifying over 90% of participants as female. One study in Sala et al. (2016)'s review and three studies in Brockmeyer et al. (2018)'s review reported 100% of participants as female. Four studies in Denison-Day et al. (2018)'s review and one study in Dahlgren and Rø (2014)'s review reported ≥98% of participants as female. Three of the reviews did not report the gender of participants in any of the studies (Isserlin et al., 2020; Tchanturia et al., 2017; Tchanturia et al., 2014). Most studies focused predominantly on female samples, therefore limiting any conclusions that could be made in terms of possible gender differences and the effectiveness of CRT. None of the reviews reported on ethnicity or socioeconomic status (SES).

3.5.2 Diagnostic status. Six of the reviews included only studies with participants with a diagnosis of AN, including AN-R (restrictive-subtype), AN-B (binge-eating/purging subtype) and atypical AN (Couturier et al., 2020; Gramaglia et al., 2020; Isserlin et al., 2020; Sala et al., 2016; Tchanturia et al., 2017; Tchanturia et al., 2014). Brockmeyer et al. (2018) included studies with participants with a former AN diagnosis and a healthy weight, participants with an ED diagnosis and a body mass index (BMI) <19, as well as participants with AN or obsessive-compulsive disorder (OCD). Reviews also included ED not otherwise specified (EDNOS) and AN, as well as AN or BN (Dahlgren & Rø, 2014; Denison-Day et al., 2018; Hagan et al., 2020). Although the current meta-review specifically focused on AN, it was noticed that there was limited CRT research in terms of other ED diagnoses. As such, due to the heterogeneity of included studies amongst the reviews, it was not possible to compare the effectiveness of CRT across ED diagnoses, or indeed specific AN subtypes.

3.5.3 Intervention and setting/format. Half of the reviews included studies conducted in outpatient and/or inpatient settings. Three of the reviews reported only inpatient settings (Gramaglia et al., 2020; Isserlin et al., 2020; Sala et al., 2016). The proportion of inpatient and/or outpatient settings was not explicitly stated in reviews by Denison-Day et al. (2018) and Tchanturia et al. (2017). Although the reviews that only reported inpatient settings questioned the overall effectiveness of CRT, the quality of these reviews were all rated as critically low, thus limiting any conclusions drawn.

Most of the reviews included studies that had CRT as an adjunct to other therapies. These included cognitive behavioural therapy (CBT), TAU, family-based treatment (FBT), behaviour therapy, multimodal inpatient treatment, and a multidisciplinary (MDT) treatment programme. Cognitive remediation was also an adjunct to emotion skills training (i.e., CREST) in Gramaglia et al. (2020) and Sala et al. (2016)'s reviews. CRT was considered a stand-alone intervention in studies included in four of the reviews (Denison-Day et al., 2018; Hagan et al., 2020; Tchanturia et al., 2017). Formats noted in terms of the delivery of CRT included: individual; group; family; computer-assisted; and manual-based. Due to variability in reporting, it was not possible to draw any conclusions in terms of the delivery of CRT (i.e., format/number of sessions) and its effectiveness. Indeed, it was difficult to decipher if the potential effectiveness of CRT was due to the intervention itself, or the existing adjunctive AN treatment. This includes RCTs included in the reviews, which typically included treatment as TAU as the control group.

Tchanturia et al. (2017) specifically referred to CRT as *not* a stand-alone treatment. As such, it is unrealistic to change complex illness-related behaviours and functioning over 8-10 sessions. Adjunctive variations include CREST, as highlighted in reviews by Gramaglia et al. (2020) and Sala et al. (2016). Preliminary support is suggested for the effectiveness of CREST as an initial intervention for AN, including in reducing alexithymia when offered in an individual format. However, it is unclear whether cognitive remediation in combination with

emotion skills training is more efficacious than other AN treatment (e.g., CBT). The quality rating for both Gramaglia et al. (2020) and Sala et al. (2016) were also critically low.

3.5.4 Comparison groups. There was a variety within and between the reviews in terms of comparison and/or control groups. Most of the reviews reported TAU as a control group, aside from Gramaglia et al. (2020) which did not report control groups at all. Three of the reviews also included studies that did not report on control and/or comparison groups (Couturier et al., 2020; Dahlgren & Rø, 2014; Tchanturia et al., 2014). Four of the reviews included studies that did not have control groups (Couturier et al., 2020; Isserlin et al., 2020; Sala et al., 2016; Tchanturia et al., 2017). This questions the quality of the studies included in the reviews and thus limiting any conclusions that can be drawn. Indeed, the inclusion of case studies within the reviews may reflect CRT being a relatively new intervention for AN (and subsequently a limited availability of existing research).

Non-specific neurocognitive therapy (NNT) was reported as a comparison group in four of the reviews (Brockmeyer et al., 2018; Dahlgren & Rø, 2014; Hagan et al., 2020; Tchanturia et al., 2014). Aside from Brockmeyer et al. (2018), these reviews also included exposure and response prevention for AN (AN-EXRP). In terms of NNT, the reviews typically reported that both groups showed high treatment acceptance, although CRT was superior to NNT in measures of cognitive set-shifting. Tchanturia et al. (2014) also notes that the authors of one of their included reviews argue that specific tailored neurocognitive training is more effective (Brockmeyer et al., 2014). The same study using AN-EXRP as a control was reported across the reviews (see Steinglass et al., 2014). It was found that AN-EXRP yielded better evidence of superior caloric intake and improvements in eating-related anxiety when compared to CRT. Again, all these reviews were rated critically low, and conclusions should be interpreted with caution.

CBT was included as a comparison group in three of the reviews (Brockmeyer et al., 2018; Hagan et al., 2020; Tchanturia et al., 2014) and with variable session numbers reported. There was found to be a lower drop-out rate and greater improvement in cognitive set-shifting

and central coherence in CRT as an adjunct to CBT versus CBT alone. Other comparison groups included art therapy (Brockmeyer et al., 2018; Couturier et al., 2020; Hagan et al., 2020) and family-based treatment (FBT; Brockmeyer et al., 2018; Hagan et al., 2020); however, there was only one published study across the reviews referencing these (Lock et al., 2018). This study showed that CRT adjuvant to FBT was associated with fewer dropouts than art therapy adjuvant to FBT (Lock et al., 2018). Denison-Day et al. (2018) also reported comparisons made between low and high autism scores as a control group in one of their studies. Indeed, the increased heterogeneity due to variability in control groups, as well as low sample sizes in the individual studies, meant that conclusions were limited.

3.5.5 Outcome measures. Although there was overlap between the reviews in terms of outcomes measures used within the studies, there was also significant variation. The outcome measures were typically able to be categorised into the following: engagement and motivation; emotional psychopathology; neuropsychological testing; and ED-related outcomes, such as increased weight.

Denison-Day et al. (2018) only reported on outcome measures relating to motivation to change, which included the motivational ruler. The only reviews that did not include outcomes related to motivation and engagement were Hagan et al. (2020) and Sala et al. (2016). Other outcome measures related to engagement and motivation included: drop-out rate (Brockmeyer et al., 2018; Tchanturia et al., 2014); the motivational stages of change for adolescents recovering from an ED (MSCARED; Couturier et al., 2020); and treatment acceptance (Dahlgren & Rø, 2014). Five of the reviews described the potential of CRT in relation to improvements in motivation to change, promoting retention, lowering dop-out rates, as well as reductions in attrition and enhancing the efficacy of concurrent treatment (Couturier et al., 2020; Dahlgren & Rø, 2014; Denison-Day et al., 2018; Tchanturia et al., 2014). However, concerns were again raised with regards to study design. This includes difficulties in being able to differentiate the effect of inpatient treatment alone, from inpatient treatment in combination with CRT (Couturier et al., 2020). As well, although Denison-Day et al. (2018) suggest that

CRT can lead to significant improvements in motivation to change, it failed to comment on the impact of sample sizes in the included studies. Whilst the review had the fewest critical and non-critical items on quality rating in comparison to the other studies, it was still rated low in terms of quality. It was therefore difficult to make any definitive conclusions in terms of engagement, motivation, and CRT.

Most of the reviews included studies using emotional psychopathology as an outcome measure. Depression was the most reported (N = 4), with specified measures including the depression inventory for children and adolescents (DIKJ) and the beck depression inventory (BDI). A study in Couturier et al. (2020)'s review found no difference between CRT and art therapy with respect to high depression scores. Dahlgren and Rø (2014) identified three studies (two based on the same sample) that reported a decrease in depression following CRT. In contrast, a meta-analysis of RCTs of CRT for AN found no effect of CRT on depressive symptoms at end of treatment. Whilst it is difficult to draw any conclusions in terms of CRT for depressive symptoms, Tchanturia et al. (2014) suggests that motivation to change may be of particular relevance for AN and depression. Other outcome measures were also concerned with social anhedonia, alexithymia, anxiety, emotion regulation, and emotion recognition. Gramaglia et al. (2020) also reported the autism spectrum quotient-10 items (AQ-10). Again, it was difficult to draw any conclusions due to the heterogeneity of included studies and quality of the reviews.

It was suggested within the reviews that studies indicate promising results in terms of neuropsychological and self-report measures (Tchanturia et al., 2017; Tchanturia et al., 2014). Eight of the reviews included outcomes related to neuropsychological testing. Examples of these include: cognitive set-shifting; central coherence; cognitive flexibility; visuospatial memory; global information processing; and verbal fluency. Three of the reviews specified the measures, which (amongst others) included: the Brixton and CatBat tasks (cognitive performance; Dahlgren & Rø, 2014); the delis–kaplan executive function system (D-KEFS); and the behaviour rating inventory of executive function (BRIEF-SR; Hagan et al., 2020;

Tchanturia et al., 2017). There is some evidence to suggest improvements in central coherence following CRT for AN, although only a small and statistically non-significant effect shown (Hagan et al., 2020; Tchanturia et al., 2017). Whilst Dahlgren and Rø (2014) and Tchanturia et al. (2014) reported that RCTs have suggested that CRT can improve cognitive set-shifting, this contrasts findings from later reviews. For example, Hagan et al. (2020)'s described variability in the measurement of set-shifting, meaning that effects cannot be determined.

Most of the reviews included outcomes specifically relevant for an ED population; although these ranged from those related to weight and eating (e.g., body mass index, calorie intake) to more direct measures (e.g., the Yale-Brown-Cornell eating disorders scale). Half of the reviews reported the ED examination questionnaire (EDE-Q) as an outcome measure, and four of the reviews identified the ED-related quality of life (QoL). Other ED-related outcomes included full remission from AN and improvement in eating-related anxiety. Several of the reviews identified promising findings for specific outcomes for AN, such as weight increase reports (Couturier et al., 2020); however, study designs meant that it was not possible to determine whether change occurred due to CRT alone. As well, no change on the EDE-Q score was reported from CRT, or improvements in ED-related symptoms or other psychological symptoms (Couturier et al., 2020; Hagan et al., 2020).

4. Discussion

The current meta-review has provided a synthesis of the relevant literature pertaining to CRT as an intervention for AN. It offers an overview of the key characteristics and findings across ten systematic reviews of varying quality. The overall quality of the included reviews according to the AMSTAR-2 framework was critically low (Shea et al., 2017), with the two reviews with the highest quality rated as low (Denison-Day et al., 2018; Tchanturia et al., 2017). This suggests that any conclusions drawn should be interpreted with caution.

The following provides an overview and discussion of the overall findings, with particular consideration towards the two reviews of better quality. Clinical implications, limitations and future directions are detailed.

4.1 Overview and discussion of findings

People with EDs are often considered to have poor motivation to change ED-related behaviours, particularly for those individuals diagnosed with AN or bulimia nervosa (BN; Leavey et al., 2011). Whilst high non-attendance rates are not unique to ED services and are recognised across NHS specialties, drop-out rates and failure to attend are amongst the highest in ED services (Leavey et al., 2011). There are a number of reasons why an individual may feel unable to take up specialist care (e.g., fear of loss of control; Eivors et al., 2003); however, it is reasonable to assume that many of those unable to access support are still in need of appropriate treatment. It is important to recognise that delays to accessing services might lead to an exacerbation and deepening of symptoms, as well as an increase in social difficulties (Leavey et al., 2011). Motivation to change is thus an important consideration within the treatment of AN.

The majority of reviews included in the current meta-review included CRT outcomes related to motivation and engagement. Central to this was Denison-Day et al. (2018)'s paper,

which specifically considered the effectiveness across a variety of therapies at improving motivation to change amongst individuals with ED. The included studies generally found significant improvements in ability but not importance to change for individuals with AN following CRT (Kuge et al., 2017; Tchanturia, Larsson, & Brown, 2016). CRT was described by Denison-Day et al. (2018) as a successful intervention with regards to this specific outcome; however, there was no consensus as to whether CRT was superior to other therapies. Whilst enhancing motivation can be considered an essential aspect of treatment, other existing NICE (2017) recommended therapies also contain motivational components (e.g., CBT; MANTRA).

An interesting outcome in one of the studies in Denison-Day et al. (2018)'s review was the exploration of any difference in change in relation to high and low levels of autistic traits. The inclusion of autism traits as an outcome was otherwise not present in any of the other reviews, aside from the autism spectrum quotient in Gramaglia et al. (2020)'s review (although no findings or conclusions drawn). Tchanturia, Larsson and Adamson (2016) found that group format CRT improves self-reported cognitive and motivational aspects in people with AN without autistic traits; however, no effects were found for those with high autistic traits. Whilst the true prevalence of autism and AN is difficult to determine and may be over-represented (see Westwood & Tchanturia, 2017), the findings of the study suggest particular consideration should be given towards individual presentation and CRT mode of delivery (e.g., more tailored CRT should be explored; Tchanturia, Larsson and Adamson, 2016). Indeed, Dinkler et al. (2021) presents the possibility that elevated autistic traits are not always present prior to the onset of AN (alternatively, coping strategies in females could mask the condition in childhood). Investigating whether autistic traits in AN are best conceptualised as a neurodevelopmental disorder (e.g., autism) or an epiphenomenon of the acute AN phase is therefore important (Dinkler et al., 2021). Regardless, the consideration of autistic traits is potentially an important aspect of CRT effectiveness.

Tchanturia et al. (2017) was another review with a quality rating of low, and thus of higher quality than the majority of reviews. The review focused on CRT for young people with AN, and again suggested that group CRT needs further investigation, with non-significant improvements found in self-reported cognitive flexibility and motivation to change (Tchanturia et al., 2017). Methodological differences across the studies made it difficult to compare findings, with studies frequently using different outcomes measures (Tchanturia et al., 2017). This difference was despite a greater focus on neuropsychological outcomes, in accordance with the cognitive processes that CRT seeks to target. The focus on neuropsychological outcomes somewhat differed from other reviews, which included emotional psychopathology and ED-related outcomes (and thus even greater variation). Tchanturia et al. (2017)'s meta-analysis of common outcome measures within the review indicated small effect size improvements in central coherence following CRT for young people. According to the review, this finding suggests that young people displayed a gestalt information processing style following the intervention (i.e., the person is better able to look at a 'bigger picture' context rather than a focus on detail). There was less consistency with regards to outcomes on set-shifting tasks across the studies, with fewer CRT sessions seemingly showing greater improvement (Tchanturia et al., 2017). It was suggested that such a finding could be explained by the small number of studies using the measure, alongside the small sample sizes in each study. Although CRT shows some potential towards the effectiveness of encouraging cognitive flexibility in young people, further research is needed to clarify such potential (Tchanturia et al., 2017).

An important finding within the current meta-review was the variation in terms of CRT setting and format. It is possible that CRT might be more effective when tailored to the individual (rather than group), which parallels findings within CBM-I research suggesting that training needs to be tailored to idiosyncratic concerns in order to show transfer (Hirsch et al.,

2016). Related to this is the emphasis on identification of individual differences that facilitate training so that methods can be usefully adapted, whilst also targeting those most likely to benefit (Hirsch et al., 2016). Yet, aside from tailoring CRT to the individual, it was otherwise difficult to draw any conclusions in terms of session number, format and/or setting (e.g., inpatient/outpatient). Given the large variety of outcome measures used, it is possible that the purpose of CRT differs for each study. However, there is a consistent use of motivational outcomes and a focus of CRT as an adjunct to other therapies. This finding suggests a central purpose of CRT for AN is to reduce drop-out rates and improve the overall effectiveness of treatment (Kim et al., 2018). However, including CRT as an adjunct to other therapies and without an appropriate comparison group (e.g., the same treatment without CRT) means that findings are limited. Nevertheless, a reasonable conclusion is that whilst the evidence is inconclusive in terms of its overall effectiveness, CRT should not be considered a stand-alone treatment (as per Tchanturia et al. (2017)'s recommendations).

4.2 Clinical implications and future directions

This meta-review presents an overview on the best available evidence of CRT for AN, ensuring that any overlap of studies between reviews was reported. Decisions and subsequent recommendations within healthcare should be based on all the available evidence, in order to draw reliable conclusions and support policy (Pieper et al., 2014). This process typically relies on the evidence from systematic reviews; however, the current meta-review has shown that findings and overall conclusions can vary. Inconsistencies were also found in terms of the quality of systematic reviews, alongside concerns with regards to overall quality according to the AMSTAR-2 framework (Shea et al., 2017).

The most promising findings of CRT for AN relate to improvements in central coherence, cognitive set-shifting, weight-increase reports, and motivation to change; however,

effects are modest and further research is needed. It was also unclear whether the potential effectiveness of CRT was due to the intervention itself, or an existing AN treatment. This is because CRT was often delivered as an adjunctive treatment alongside an existing therapy, thus making it difficult to disentangle the effects of CRT from the combined therapy. For example, weight restoration in AN might lead to improvements in global processing, and cognitive restructuring in CBT might result in an improvement in flexible thinking. As such, it may be premature to recommend the intervention as an adjunctive or stand-alone treatment for AN until further high-quality research is carried out. However, as described by Watson and Bulik (2013), a lack of evidence is not commensurate with 'lack of effectiveness', or indeed a 'do nothing' approach. Appropriate management and treatment for AN can save a person's life, as well as substantially improve quality of life. Uncertainty exists amongst the various treatments available for AN, including the comparative effectiveness of interventions (Watson & Bulik, 2013).

A key finding revealed within this meta-review with regards to the direction of future research is the narrowing down of outcome measures used. CRT was developed to target specific cognitive processes, and it is therefore possible that any outcome measure beyond that of neuropsychological testing should be considered secondary (e.g., emotional psychopathology; ED-related outcomes). To minimise confusion with regards to which outcomes measures to use, specific measures of set-shifting and central coherence need to be clearly defined for the use of CRT for AN research. This would allow for comparisons between CRT studies to be made. Indeed, well-designed large-scale studies of CRT are needed, including appropriate control groups allowing for CRT to be differentiated from its adjunctive treatment. As well, there are currently no studies that incorporate CBM as a comparison or adjunct to CRT, even though it might be theoretically justified to do so. CBM similarly targets cognitive processes, but more involuntary (bottom-up) ones, in contrast to voluntary (top-

down) information processes as per CRT. Indeed, there are limitations with ongoing case series, and research incorporating larger sample sizes should be a priority. It would also be helpful to explore the optimal numbers of CRT sessions in both individual and group format. Lastly, the current literature almost solely focuses on CRT for AN, with limited reference to other ED diagnoses. Yet, there is evidence that people with BN clearly have central coherence and set-shifting inefficiencies, which do not differ significantly from those observed in people with AN (Keegan et al., 2021). It might therefore be appropriate to extend the application of CRT across EDs, and that people with BN might similarly benefit from such adjunctive approaches.

4.3 Limitations

Although a meta-review pulls together the most consistent findings from the highest quality reviews available, it does not allow for the level of in-depth examination found within systematic reviews (Stott et al., 2021). It is therefore possible that the current review did not include recent primary studies that may have revealed important findings. Indeed, overviews (reviews of reviews) are a relative new type of synthesis of the evidence, having gained increased interest due to the steady increase in systematic reviews (Hartling et al., 2012). A potential advantage is that they enable comparison between conditions (or interventions); however, this was not possible within the present review due to the heterogenous nature of the included studies. Alongside this, the poor quality of the reviews meant that few conclusions could be drawn.

A further limitation of the current study is that the publication of the AMSTAR-2 was fairly recent (Shea et al., 2017); and thus its limitations yet to be revealed. Whilst identification of critical items was approached with flexibility, it should be noted that excluding item two (i.e., protocol registration) does not mean that it is not best practice. Nevertheless, given the

low ratings of the included reviews, inclusion of item two as a critical item would have made little difference to the overall conclusions. It should also be noted that the current meta-review did not itself pre-register a protocol. Indeed, registration in the international prospective register of systematic reviews (PROSPERO) is associated with increased review quality (Sideri et al., 2018). A priori registration of systematic review protocols can help to reduce selective reporting of outcomes and should be encouraged when possible (Sideri et al., 2018). Limitations within the current review also include only one reviewer having performed data extraction. This increases opportunity for error and bias reporting of results; for example, identifying whether the included reviews found CRT to be helpful overall (i.e., 'yes' or 'no') was not always clear and relied on reviewer judgment. Indeed, there are limitations with this classification approach, whereby it does not consider risk of bias in terms of the narrative synthesis provided by the authors. The classification approach may also obscure some of the nuance in terms of the findings; for example, it does not consider the different outcomes that the reviews are looking at, the mode of delivery of CRT included in the review, or indeed the strength of the papers. The approach assumes that the helpfulness or unhelpfulness of CRT in terms of clinical outcomes are equally weighted, whereas a potential strength in one area (e.g., improvements in neuropsychological testing) might overshadow another (e.g., no improvement in calorie intake).

5. Conclusions

There is an ever-growing body of literature interested in understanding the cognitive mechanisms involved in the identification and maintenance of EDs, and thus an increase in related research regarding emerging treatments. A particular increase in studies include those examining CRT for AN, with a number of published systematic reviews appraising the best

available evidence. Given the continued uncertainty pertaining to the effectiveness of CRT, this paper provides a meta-review of reviews including CRT as an intervention for AN.

The ten identified reviews broadly suggest that CRT is helpful within the treatment of AN. Promising findings relate to improvements in central coherence, cognitive set-shifting, weight-increase reports, and motivation to change. Yet, critical limitations with regards to the methodological quality of the reviews and the design of the included studies, preclude the ability to draw any definitive conclusions. With eighty percent of the reviews deemed critically low in terms of overall quality, future systematic reviews would benefit from a focus on methodological improvement. This might include considering the impact of small studies on findings, pre-registering a research protocol, as well as implementing a more detailed consideration of RoB of primary studies.

6. References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). American Psychiatric Association.

Attia, E. (2021). Anorexia nervosa treatment trials: time for new approaches. *The Lancet, 8*(3), 170-171.

Berg, K. C., & Wonderlich, S. A. (2013). Emerging psychological treatments in the field of eating disorders. *Current Psychiatry Reports, 15*(11), 407.

Brockmeyer, T., Friederich, H.-C., & Schmidt, U. (2018). Advances in the treatment of anorexia nervosa: a review of established and emerging interventions. *Psychological Medicine, 48*(8), 1228-1256.

Brockmeyer, T., Ingenerf, K., Walther, S., Wild, B., Hartmann, M., Herzog, W., Bents, H., & Friederich, H. C. (2014). Training cognitive flexibility in patients with anorexia nervosa: a pilot randomized controlled trial of cognitive remediation therapy. *International Journal of Eating Disorders, 47*(1), 24-31.

Burda, B. U., Holmer, H. K., & Norris, S. L. (2016). Limitations of A Measurement Tool to Assess Systematic Reviews (AMSTAR) and suggestions for improvement. *Systematic Reviews, 5*(1), 1-10.

Caglar-Nazali, H. P., Corfield, F., Cardi, V., Ambwani, S., Leppanen, J., Olabintan, O., Deriziotis, S., Hadjimichalis, A., Scognamiglio, P., & Eshkevari, E. (2014). A systematic review and meta-analysis of 'Systems for Social Processes' in eating disorders. *Neuroscience & Biobehavioral Reviews, 42*, 55-92.

Cicerone, K. D., Langenbahn, D. M., Braden, C., Malec, J. F., Kalmar, K., Fraas, M., Felicetti, T., Laatsch, L., Harley, J. P., & Bergquist, T. (2011). Evidence-based

cognitive rehabilitation: updated review of the literature from 2003 through 2008. *Archives of Physical Medicine and Rehabilitation, 92*(4), 519-530.

Cooper, M. J. (2005). Cognitive theory in anorexia nervosa and bulimia nervosa: Progress, development and future directions. *Clinical Psychology Review, 25*(4), 511-531.

Couturier, J., Isserlin, L., Norris, M., Spettigue, W., Brouwers, M., Kimber, M., McVey, G., Webb, C., Findlay, S., & Bhatnagar, N. (2020). Canadian practice guidelines for the treatment of children and adolescents with eating disorders. *Journal of Eating Disorders, 8*(1), 1-80.

Dahlgren, C. L., & Rø, Ø. (2014). A systematic review of cognitive remediation therapy for anorexia nervosa-development, current state and implications for future research and clinical practice. *Journal of Eating Disorders, 2*(1), 1-12.

Denison-Day, J., Appleton, K. M., Newell, C., & Muir, S. (2018). Improving motivation to change amongst individuals with eating disorders: A systematic review. *International Journal of Eating Disorders, 51*(9), 1033-1050.

Dinkler, L., Taylor, M. J., Råstam, M., Hadjikhani, N., Bulik, C. M., Lichtenstein, P., Gillberg, C., & Lundström, S. (2021). Anorexia nervosa and autism: a prospective twin cohort study. *Journal of Child Psychology and Psychiatry, 62*(3), 316-326.

Eichen, D. M., Matheson, B. E., Appleton-Knapp, S. L., & Boutelle, K. N. (2017). Neurocognitive treatments for eating disorders and obesity. *Current Psychiatry Reports, 19*(9), 1-10.

Eivors, A., Button, E., Warner, S., & Turner, K. (2003). Understanding the experience of drop-out from treatment for anorexia nervosa. *European Eating Disorders Review: The Professional Journal of the Eating Disorders Association, 11*(2), 90-107.

Faggion, C. M. (2015). Critical appraisal of AMSTAR: challenges, limitations, and potential solutions from the perspective of an assessor. *BMC Medical Research Methodology, 15*(1), 1-5.

Galmiche, M., Déchelotte, P., Lambert, G., & Tavolacci, M. P. (2019). Prevalence of eating disorders over the 2000–2018 period: a systematic literature review. *The American Journal of Clinical Nutrition, 109*(5), 1402-1413.

Gramaglia, C., Gambaro, E., & Zeppegno, P. (2020). Alexithymia and Treatment Outcome in Anorexia Nervosa: A Scoping Review of the Literature. *Frontiers in Psychiatry, 10*, 1-10.

Green, M. W., Elliman, N. A., Wakeling, A., & Rogers, P. J. (1996). Cognitive functioning, weight change and therapy in anorexia nervosa. *Journal of Psychiatric Research, 30*(5), 401-410.

Hagan, K. E., Christensen, K. A., & Forbush, K. T. (2020). A preliminary systematic review and meta-analysis of randomized-controlled trials of cognitive remediation therapy for anorexia nervosa. *Eating Behaviors, 37*, 101391.

Hartling, L., Chisholm, A., Thomson, D., & Dryden, D. M. (2012). A descriptive analysis of overviews of reviews published between 2000 and 2011. *PloS one, 7*(11), e49667.

Hay, P. (2020). Current approach to eating disorders: a clinical update. *Internal Medicine Journal, 50*(1), 24-29.

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.

Isserlin, L., Spettigue, W., Norris, M., & Couturier, J. (2020). Outcomes of inpatient psychological treatments for children and adolescents with eating disorders at time of discharge: a systematic review. *Journal of Eating Disorders, 8*(1), 1-17.

Jüni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association, 282*(11), 1054-1060.

Keegan, E., Tchanturia, K., & Wade, T. D. (2021). Central coherence and set-shifting between nonunderweight eating disorders and anorexia nervosa: A systematic review and meta-analysis. *International Journal of Eating Disorders, 54*(3), 229-243.

Khalsa, S. S., Portnoff, L. C., McCurdy-McKinnon, D., & Feusner, J. D. (2017). What happens after treatment? A systematic review of relapse, remission, and recovery in anorexia nervosa. *Journal of Eating Disorders, 5*(1), 1-12.

Kim, E. J., Bahk, Y.-C., Oh, H., Lee, W.-H., Lee, J.-S., & Choi, K.-H. (2018). Current status of cognitive remediation for psychiatric disorders: a review. *Frontiers in Psychiatry, 9*, 461.

Kothari, R., Solmi, F., Treasure, J., & Micali, N. (2013). The neuropsychological profile of children at high risk of developing an eating disorder. *Psychological Medicine, 43*(7), 1543-1554.

Kuge, R., Lang, K., Yokota, A., Kodama, S., Morino, Y., Nakazato, M., & Shimizu, E. (2017). Group cognitive remediation therapy for younger adolescents with anorexia nervosa: a feasibility study in a Japanese sample. *BMC Research Notes, 10*(1), 1-7.

Lang, K., Lopez, C., Stahl, D., Tchanturia, K., & Treasure, J. (2014). Central coherence in eating disorders: an updated systematic review and meta-analysis. *The World Journal of Biological Psychiatry, 15*(8), 586-598.

Leavey, G., Vallianatou, C., Johnson-Sabine, E., Rae, S., & Gunputh, V. (2011). Psychosocial barriers to engagement with an eating disorder service: A qualitative analysis of failure to attend. *Eating Disorders, 19*(5), 425-440.

Lena, S. M., Fiocco, A. J., & Leyenaar, J. K. (2004). The role of cognitive deficits in the development of eating disorders. *Neuropsychology Review, 14*(2), 99-113.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., 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.

Lock, J., Fitzpatrick, K. K., Agras, W. S., Weinbach, N., & Jo, B. (2018). Feasibility study combining art therapy or cognitive remediation therapy with family-based treatment for adolescent anorexia nervosa. *European Eating Disorders Review, 26*(1), 62-68.

Lopez, C., Davies, H., & Tchanturia, K. (2012). Neuropsychological inefficiences in anorexia nervosa targeted in clinical practice: the development of a module of cognitive remediation therapy. In J. Fox & K. Goss (Eds.), *Eating and its Disorders*. John Wiley & Sons.

Matheson, E., Wade, T. D., & Yiend, J. (2019). Utilising cognitive bias modification to remedy appearance and self-worth biases in eating disorder psychopathology: A systematic review. *Journal of Behavior Therapy and Experimental Psychiatry, 65*, 101482.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine, 6*(7), 1-6.

Mulrow, C. D. (1994). Systematic reviews: rationale for systematic reviews. *British Medical Journal, 309*(6954), 597-599.

NICE. (2017). *Eating disorders: recognition and treatment*. National Institute for Health and Care Excellence Guideline (NG69).

Pieper, D., Antoine, S.-L., Mathes, T., Neugebauer, E. A., & Eikermann, M. (2014). Systematic review finds overlapping reviews were not mentioned in every other overview. *Journal of Clinical Epidemiology, 67*(4), 368-375.

Pieper, D., Puljak, L., Lorenzo, G., & Minozzi, S. (2018). Comparison of AMSTAR 2 with ROBIS in systematic reviews including randomized and non-randomized studies. Abstracts of the 25th Cochrane colloquium, Edinburgh, UK. *Cochrane Database of Systematic Reviews, 9*(Suppl 1).

Pollock, M., Fernandes, R. M., Becker, L. A., Pieper, D., & Hartling, L. (2018). Chapter V: overviews of reviews. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (Vol. 6). John Wiley & Sons.

Roberts, M. E., Tchanturia, K., & Treasure, J. L. (2013). Is attention to detail a similarly strong candidate endophenotype for anorexia nervosa and bulimia nervosa? *The World Journal of Biological Psychiatry, 14*(6), 452-463.

Rowlands, K., Wilson, E., Simic, M., Harrison, A., & Cardi, V. (2020). A Critical Review of Studies Assessing Interpretation Bias Towards Social Stimuli in People With Eating Disorders and the Development and Pilot Testing of Novel Stimuli for a Cognitive Bias Modification Training. *Frontiers in Psychology, 11*, 538527.

Sala, M., Heard, A., & Black, E. A. (2016). Emotion-focused treatments for anorexia nervosa: a systematic review of the literature. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 21*(2), 147-164.

Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., Henry, D. A., & Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology, 62*(10), 1013-1020.

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., & Kristjansson, E. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *British Medical Journal, 358*, 1-9.

Sideri, S., Papageorgiou, S. N., & Eliades, T. (2018). Registration in the international prospective register of systematic reviews (PROSPERO) of systematic review protocols was associated with increased review quality. *Journal of Clinical Epidemiology, 100*, 103-110.

Smink, F. R., Van Hoeken, D., & Hoek, H. W. (2012). Epidemiology of eating disorders: incidence, prevalence and mortality rates. *Current Psychiatry Reports, 14*(4), 406-414.

Smith, K. E., Mason, T. B., Johnson, J. S., Lavender, J. M., & Wonderlich, S. A. (2018). A systematic review of reviews of neurocognitive functioning in eating disorders: The state-of-the-literature and future directions. *International Journal of Eating Disorders, 51*(8), 798-821.

Smith, V., Devane, D., Begley, C. M., & Clarke, M. (2011). Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Medical Research Methodology, 11*(1), 1-6.

Solmi, M., Wade, T., Byrne, S., Del Giovane, C., Fairburn, C., Ostinelli, E., De Crescenzo, F., Johnson, C., Schmidt, U., & Treasure, J. (2021). Comparative efficacy and acceptability of psychological interventions for the treatment of adult outpatients with anorexia nervosa: a systematic review and network meta-analysis. *The Lancet Psychiatry, 8*, 215-224.

Southgate, L., Tchanturia, K., & Treasure, J. (2005). Building a model of the aetiology of eating disorders by translating experimental neuroscience into clinical practice. *Journal of Mental Health, 14*(6), 553-566.

Steinglass, J. E., Albano, A. M., Simpson, H. B., Wang, Y., Zou, J., Attia, E., & Walsh, B. T. (2014). Confronting fear using exposure and response prevention for anorexia nervosa: A randomized controlled pilot study. *International Journal of Eating Disorders, 47*(2), 174-180.

Stott, N., Fox, J. R., & Williams, M. O. (2021). Attentional bias in eating disorders: A meta-review. *International Journal of Eating Disorders, 54*, 1377-1399.

Stuhldreher, N., Konnopka, A., Wild, B., Herzog, W., Zipfel, S., Löwe, B., & König, H. H. (2012). Cost-of-illness studies and cost-effectiveness analyses in eating disorders: A systematic review. *International Journal of Eating Disorders, 45*(4), 476-491.

Sullivan, P. F. (1995). Mortality in anorexia nervosa. *American Journal of Psychiatry, 152*(7), 1073-1074.

Szmukler, G. I., Andrewes, D., Kingston, K., Chen, L., Stargatt, R., & Stanley, R. (1992). Neuropsychological impairment in anorexia nervosa: before and after refeeding. *Journal of Clinical and Experimental Neuropsychology, 14*(2), 347-352.

Tchanturia, K., Davies, H., & Campbell, I. C. (2007). Cognitive remediation therapy for patients with anorexia nervosa: preliminary findings. *Annals of General Psychiatry, 6*(1), 1-6.

Tchanturia, K., Davies, H., Reeder, C., & Wykes, T. (2010). *Cognitive remediation therapy for anorexia nervosa*. King's College London, University of London.

Tchanturia, K., Davies, H., Roberts, M., Harrison, A., Nakazato, M., Schmidt, U., Treasure, J., & Morris, R. (2012). Poor cognitive flexibility in eating disorders: examining the evidence using the Wisconsin Card Sorting Task. *PloS One, 7*(1), e28331.

Tchanturia, K., Giombini, L., Leppanen, J., & Kinnaird, E. (2017). Evidence for cognitive remediation therapy in young people with anorexia nervosa: Systematic review and meta-analysis of the literature. *European Eating Disorders Review, 25*(4), 227-236.

Tchanturia, K., Larsson, E., & Adamson, J. (2016). How anorexia nervosa patients with high and low autistic traits respond to group Cognitive Remediation Therapy. *BMC Psychiatry, 16*(1), 1-7.

Tchanturia, K., Larsson, E., & Brown, A. (2016). Benefits of group cognitive remediation therapy in anorexia nervosa: case series. *Neuropsychiatrie, 30*(1), 42-49.

Tchanturia, K., Lounes, N., & Holttum, S. (2014). Cognitive remediation in anorexia nervosa and related conditions: a systematic review. *European Eating Disorders Review, 22*(6), 454-462.

Treasure, J. L. (2007). Getting beneath the phenotype of anorexia nervosa: the search for viable endophenotypes and genotypes. *The Canadian Journal of Psychiatry, 52*(4), 212-219.

Turton, R., Cardi, V., Treasure, J., & Hirsch, C. R. (2018). Modifying a negative interpretation bias for ambiguous social scenarios that depict the risk of rejection in women with anorexia nervosa. *Journal of Affective Disorders, 227*, 705-712.

Udo, T., & Grilo, C. M. (2018). Prevalence and correlates of DSM-5–defined eating disorders in a nationally representative sample of US adults. *Biological Psychiatry, 84*(5), 345-354.

Ulfvebrand, S., Birgegård, A., Norring, C., Högdahl, L., & von Hausswolff-Juhlin, Y. (2015). Psychiatric comorbidity in women and men with eating disorders results from a large clinical database. *Psychiatry Research, 23*(2), 294-299.

Waller, G. (2016). Recent advances in psychological therapies for eating disorders. *F1000Research, 5*(F1000 Faculty Rev), 702.

Watson, H., & Bulik, C. (2013). Update on the treatment of anorexia nervosa: review of clinical trials, practice guidelines and emerging interventions. *Psychological Medicine, 43*(12), 2477-2500.

Westwood, H., & Tchanturia, K. (2017). Autism spectrum disorder in anorexia nervosa: an updated literature review. *Current Psychiatry Reports, 19*(7), 41.

Wykes, T., Huddy, V., Cellard, C., McGurk, S. R., & Czobor, P. (2011). A meta-analysis of cognitive remediation for schizophrenia: methodology and effect sizes. *American Journal of Psychiatry, 168*(5), 472-485.

Zipfel, S., Giel, K. E., Bulik, C. M., Hay, P., & Schmidt, U. (2015). Anorexia nervosa: aetiology, assessment, and treatment. *The lancet psychiatry, 2*(12), 1099-1111.

Chapter Two: Empirical paper

Developing a measure of interpretation bias for worry:

a study of reliability and validity

Word count: 12,740

(exc. Abstract and references)

Article intended for submission to the Journal of Anxiety Disorders for peer review

Highlights

- A worry-related version of the Word Sentence Association Paradigm (WSAP) as a measure of interpretation bias was developed.
- The developed measure yielded positive results with regards to its psychometric properties, including performance over time.
- Trait worry was found to be closely related to the developed worry-related interpretation bias measure.
- There is evidence for the association of worry and interpretation bias.

Abstract

Interpretation bias (IB) refers to the tendency to draw inferences about information in consistent ways and is a causal process across various mental health difficulties. It is thought that negative IB is a key mechanism underlying excessive worry, which is a cardinal feature of generalised anxiety disorder (GAD). Further experimental research regarding the underlying mechanisms involved in both the generation and modification of IB is needed. Yet, there is a key limitation in the current literature on IB, whereby there is a need to test the reliability and validity of the methods used.

This study has sought to build on existing IB methodologies by psychometrically evaluating a version of the Word Sentence Association Paradigm (WSAP). The objective of the study was to adapt and develop existing material (i.e., a list of ambiguous statements) for a worry-related IB task, to be used for assessment and training purposes. The newly developed measure has been referred to as the Statement Word Association Task (SWAT). Participants were presented with a series of (grammatically complete) ambiguous statements and asked to resolve the ambiguity of each statement with a single word. Each item was reviewed by members of the research team, and we sought to validate the measure by rating the resolutions as either positive or negative in valence. The range of responses and the emotional valence generated were explored. The new measure of IB was evaluated against a number of existing measures related to IB, anxiety, worry and depression.

The measure yielded positive results with regards to its psychometric properties. Included within the study are findings related to content validity, criterion validity, convergent validity, test-retest reliability, inter-rater reliability, and internal consistency. The key finding was that trait worry is related to performance on the SWAT. As well, hierarchical regression analysis showed that worry made the strongest unique contribution to explaining the scores on

the SWAT, providing evidence for the association of worry and IB. An overview and discussion of the findings, clinical implications, limitations, and future directions are detailed.

Keywords:

Worry; Interpretation bias*;* Measurement of interpretation bias*;* Reliability; Validity

1. Introduction

Worry is a common experience, and everybody worries from time to time (Ruscio, 2002). It can be understood as a chain of future-oriented thoughts and images involving uncertain and possibly negative outcomes (Sibrava & Borkovec, 2006). This process of worry represents an attempt to engage in mental problem solving of an issue (Borkovec et al., 1983). It is typically experienced as mild, transient, and limited in scope (Ruscio, 2002). Pathological worry is the term used when worry is considered uncontrollable, related closely to fear processes, and is negative in valence (Borkovec et al., 1983).

Excessive and uncontrollable worry about a number of future events or activities is a cardinal defining feature of generalised anxiety disorder (GAD; American Psychiatric Association, 2013). It is associated with a number of difficulties (with or without a GAD diagnosis), including stress, poor perceived physical health, as well as the avoidance of processing negative emotions (Kertz & Woodruff-Borden, 2011; Llera & Newman, 2010; Szabó, 2011). There is also evidence of a relationship between worry, sleep disturbance, and depressive symptoms (Xu et al., 2020). Uncontrollable worry can therefore have a significant impact on an individuals' overall health, adaptive functioning, and wellbeing. This is an important consideration given that many people (28%) experience uncontrollable worry, whilst only a minority of people (6%) meet criteria for GAD (Ruscio, 2002).

1.1 Theoretical models of worry

The importance of worry as a feature of GAD has been emphasised by various theoretical models within the literature. For example, Wells (1995) presents a meta-cognitive classification and analysis of factors that contribute to the development of worry. The model proposes that problematic worry is maintained by maladaptive metacognitions (e.g., negative appraisal of worrying and linked behaviours; negative beliefs about worry concerning

uncontrollability and danger). There is also the intolerance of uncertainty model, which has shown to be important in the development and maintenance of worry (Dugas et al., 1998). Intolerance of uncertainty is defined by Ladouceur et al. (1997) as the way in which an individual perceives and responds to ambiguous or uncertain information with a set of cognitive, behavioural and emotional reactions. Alongside this, Borkovec et al., (2004)'s avoidance theory presents empirical support that worry is a perseverative, cognitive activity that serves as an avoidance function. There is also an emotional dysregulation model (Mennin et al., 2005), whereby individuals may utilise worry and other intra- and interpersonal processes to avoid the distress associated with heightened emotional intensity. The current study is specifically interested in models of worry focused on component processes for which there is evidence that they may have a causal role in worry, which is of importance when considering implications for treatment. Aspects of the models to be explored overlap with previous accounts of pathological worry, whereby several factors may combine to promote worry.

Hirsch and Mathews (2012) present a cognitive model concerning the onset and maintenance of pathological worry. The model proposes that worry arises from an interaction between involuntary (bottom-up) processes (e.g., biases in attention and interpretation in favour of threat content), and voluntary (top-down) processes (e.g., attentional control). For example, people vary in how they interpret ambiguous events (Hertel & Mathews, 2011). Those who experience anxiety are more likely than others to interpret ambiguity in a negative way, attend more to emotionally negative cues, and selectively recall negative information (Mathews & MacLeod, 2005). These processes are then likely to foster the generation of additional negative thoughts, that could trigger or maintain episodes of worry. The model also proposes that high worriers have poor attention control; that is, the ability to shift attention from one topic to another, or to intentionally ignore distracting information (Hirsch & Mathews, 2012; Miyake et al., 2000).

Davey and Meeten (2016) present a model of worry that examines an integrated cascading network of cognitive and behavioural reactions. The model extends Hirsch and Mathews (2012)'s information processing approach, specifying behavioural and cognitive processes downstream from attentional and interpretational bias, that contribute to the perseverative nature of worry in pathological worriers (Davey & Meeten, 2016). These include positive beliefs in worry (e.g., 'worry helps analytic thinking'), less ability to exert attention control over worry thoughts, and stringent goal-directed rules for dealing with threat (e.g., 'I must consider every possible outcome'). The role of negative mood also promotes a more systematic and deliberate information processing style, thus providing negative information for evaluating the merit of the worry (Ambady & Gray, 2002; Davey & Meeten, 2016).

In review of the models, the role of information processing biases and poor or misdirected attention control appear key towards understanding the development and maintenance of worry. It is therefore important to look at the underlying cognitive mechanisms of worry, in order to inform effective treatment and interventions. Yet, further information is needed with regards to understanding such cognitive processing factors. The current study has focused on one particular type of cognitive bias, namely interpretation bias (IB).

1.1.1 Interpretation bias and worry. IB is a common mechanism across various mental health difficulties (Hirsch et al., 2016), including GAD (e.g., Anderson et al. 2012; Mathews et al. 1989) and depression (Everaert et al. 2017). IB refers to the tendency to draw inferences about ambiguous information in consistent ways. This may be done in a benign (i.e., neutral or positive) or a negative manner (Hirsch et al., 2016). For example, a smile may suggest friendliness or agreement, but also intimidation or mockery (Hirsch et al., 2016). This will then have implications on how a person responds or reacts (e.g., either with warmth or resentment). As such, the interpretation of ambiguity may have consequences for both mood and behaviour (Hirsch et al., 2016).

Hirsch and Mathews (2012) and Davey and Meeten (2016) propose IB as a key maintaining factor of pathological worry. Resolving ambiguity is ubiquitous to everyday life, yet individuals with high levels of worry tend to make relatively threatening interpretations in the face of ambiguous information (Hirsch & Mathews, 2012; Krahé et al., 2019). An example of a negative IB might be a situation where a person is late to hear back about the results of an exam they have taken, whereby they interpret this to mean they have failed (conversely, a positive/benign interpretation might be that the marker was running late). Krahé et al. (2019) explored the relationship between IB and forms of repetitive negative thinking (i.e., worry and rumination) in clinical versus nonclinical samples. This cross-sectional study found that individuals with a diagnosis of GAD and high levels of worry generated significantly more negative interpretations than healthy comparison participants (Krahé et al., 2019). The findings thus support a relationship between IB and worry. Research by Feng et al. (2019) also found that high worriers lacked the benign IB that low worriers had throughout different stages of information processing, as shown by behavioural and neurophysiological interpretation indices. The stages of information processing include: either after reflecting on ambiguous information, or when individuals first encounter ambiguous information (Hirsch et al., 2016). IB has also been found to predict future worry and anxiety when individuals are under stress, indicating that benign IB may help to account for individual differences in stress tolerance.

It is thought that negative IB is a key mechanism underlying pathological worry (Hirsch et al., 2016). Thus, worry-related IB is an important concept to investigate, as central to the current study (Feng et al., under review).

1.2 Cognitive bias modification procedures and interpretation bias

Cognitive bias modification (CBM) procedures have been designed to test the theoretical claim that selective information processing plays an important causal role in the

development and maintenance of emotional psychopathology (Koster et al., 2009). These procedures are designed to change particular styles of cognitive processing that are thought to contribute to undesirable emotional reactions or disorders (Koster et al., 2009). Experiments utilising CBM have demonstrated that cognitive biases can indeed be modified and subsequently influence emotional reactivity (Hertel & Mathews, 2011). Cognitive bias modifications for interpretations (CBM-I) refers to an experimental training paradigm pioneered by Grey and Mathews (2000) and Mathews and Mackintosh (2000), which repeatedly resolves ambiguous situations to favour consistent (i.e., benign) interpretations. CBM-I paradigms have shown to be an effective method in training a more benign IB (Hirsch et al., 2016). This includes for individuals with high levels of worry and those with GAD (Hayes et al., 2010; Hirsch et al., 2009; Hirsch et al., 2021).

IB can be assessed using a range of different methods, which are often understood as either 'offline' or 'online' paradigms (Hirsch et al., 2016). Participants are allowed to reflect on ambiguous material in 'offline' paradigms, without having to report the first inference that comes to mind (Hirsch et al., 2016). Such research includes that of Butler and Mathews (1983), which showed that when individuals with a diagnosis of GAD ranked ambiguous scenarios, they selected threatening interpretations more highly than did non-anxious controls. As well, individuals with GAD have been found to endorse more threat-related interpretations when presented with ambiguously threatening scenarios (Eysenck et al., 1991). 'Online' paradigms require participants to make speeded responses to interpretations, as generated at the time the ambiguity is first encountered (Hirsch et al., 2016; Krahé et al., 2016). For example, Hirsch and Mathews (2000) used a speeded lexical decision task following ambiguous threatening situations (e.g., being interviewed for a job), and at critical points participants responded to targets matching possible interpretations. The study found that non-anxious participants typically made strong positive interpretations, whereas people with a diagnosis of social

anxiety disorder (SAD) did not.

There is strong evidence concerning an IB and anxiety link, suggesting a causal role for negative IB in the development and maintenance of worry in GAD (Hazlett-Stevens & Borkovec, 2004; Hirsch et al., 2018; Hirsch & Mathews, 2012; Krahé et al., 2019). This is not only of theoretical importance, but also of clinical relevance when considering IB as a promising target for developing effective interventions (Beard, 2011; Hirsch et al., 2018). NICE (2011) guidelines recommend cognitive behavioural therapy (CBT) as a treatment for GAD, and whilst it is effective in reducing symptoms, around half of those treated fail to achieve high end-state functioning (Borkovec et al., 2002). Hayes et al. (2010) discuss the focus of CBT interventions around challenging negative interpretations, suggesting interpretation modification techniques as a potentially useful addition to existing strategies.

1.3 The measurement of interpretation bias

There is a key limitation in the current literature on IB, whereby there is a need to test the reliability and validity of the methods used. This is important for the advancement of research on IB in emotional disorders, including those studies looking to assess the causal role of IB in maintaining central components of anxiety using CBM-I (Hirsch et al., 2016). Whilst a number of measures have been developed to assess and/or train a certain IB (see Hirsch et al. (2016) for further detailed examples), often the measures used have not been assessed for reproducibility and/or meaning of construct. This is crucial to ensure quality for both IB training and assessment purposes. Thus, prior to the development of studies seeking to 'train' a certain interpretative style, it is important to look at whether the measures used are indeed assessing the psychological concept under investigation.

Psychometric theory is the theory of psychological measurement and offers a framework for evaluating specific measurement instruments (Barker et al., 2016). The central

concepts are reliability and validity. Reliability concerns the reproducibility of measurement and the degree to which the measure controls for random error (Barker et al., 2016). Reasonably high reliability enables measurement with precision, allowing relationships between variables to be explored. The principal subtypes of reliability include test-retest reliability, equivalent forms, internal consistency (including split-half), and inter-rater reliability (Barker et al., 2016). The process of developing and evaluating measures involves a two-step process, whereby for a measure to be valid, it must first be reliable. Validity is used to assess the meaning of a measurement, that is, the extent to which a test measures what it purports to measure (Colman, 2006). It can be divided into content, face, criterion, and construct validity, and a thoroughly researched measure will report all of these (Barker et al., 2016).

A systematic review identified forty-five experimental paradigms available for exploring IB across a range of emotional disorders, which have been designed and implemented (Schoth & Liossi, 2017). The paradigms were categorised into those using ambiguous words, ambiguous images, and ambiguous scenarios. The review highlights the importance of the validation of the experimental stimuli to be used within the paradigms, in order for IB to be reliably explored. However, it does not detail any studies which have validated their tasks, despite this being deemed an essential consideration prior to running any investigation (Schoth & Liossi, 2017). An identified limitation of certain studies is the lack of disorder-related content used within the paradigm, thus questioning the specificity of bias being explored (In-Albon et al., 2008; Rohrbacher & Reinecke, 2014).

One of the paradigms included in Schoth and Liossi (2017)'s review is the Word Sentence Association Paradigm (WSAP). The original version of the WSAP involves presenting a single prime word (benign or threatening) followed by an ambiguous sentence, whereby participants indicate whether the two are related or not (Beard & Amir, 2009). A

systematic review by Gonsalves et al. (2019) found that the WSAP has been used to target 18 disorders and symptoms in adults and children. Strengths of the paradigm include flexibility in exploring the impact of different categories (e.g., negative, positive, disorder-specific) of primes and target sentences (Gonsalves et al., 2019). Adequate reliability and validity of the WSAP indices was reported for adults with elevated symptoms of anxiety, hostility, and depression (Gonsalves et al., 2019). However, most of the reliability estimates were obtained from a specific version of the paradigm (i.e., WSAP-Hostility, Likert scale). Alongside this, only twelve of the forty-two studies reported reliability estimates, and only two studies included test-retest reliability (Dillon et al., 2016; Martinelli et al., 2014).

There have been many adaptations of the WSAP within IB research; yet, many studies that changed the task did not mention the changes within their methods sections (rather presenting their procedure as though it were standard; Gonsalves et al. 2019). As such, it is important that future studies clearly state any modifications to the original WSAP, including a rationale for the change. In terms of worry-related IB, there was only one study identified within Gonsalves et al. (2019)'s review with this specific interest. Ogniewicz et al. (2014) adapted the WSAP in order to assess IB with regards to various domains of life that are typical among individuals with GAD and pathological worry. Participants were asked to accept or reject threat and benign interpretations of ambiguous sentences. The results provide evidence for the convergent validity of the WSAP for generalised anxiety. However, limitations include participants having only been tested on one single occasion, therefore the reliability of the measure over time could not be established. Indeed, reliability of the WSAP over time is crucial for studies using the outcome measure before and after interventions (e.g., CBM-I; Gonsalves et al. 2019). In addition, participants may have responded to particular trials without having understood the antecedent words, due to them being presented only very briefly (for 500ms) on a computer monitor (Ogniewicz et al., 2014).

1.4 The current study

In order to improve effectiveness and efficiency of psychological interventions, focus should be given towards gaining a better understanding of the mechanisms that maintain the cardinal feature of GAD; that is, excessive worry (Hayes et al., 2010). Further experimental research regarding the underlying mechanisms involved in both the generation and modification of IB is therefore needed (Hirsch et al., 2016). The selection of a specific paradigm to assess IB in emotional disorders should be determined by the research question, although utilising a range of paradigms is likely to be most robust (Hirsch et al., 2016). However, measures and materials need to be evaluated in order to assess whether the training paradigms are doing what they are meant to be doing. The current study has therefore sought to build on existing IB methodologies by psychometrically evaluating a version of the WSAP measure of IB, specifically in relation to worry.

The objective of the study was to adapt and develop existing material (i.e., a list of ambiguous statements) for a worry-related version of the WSAP task, to be used for assessment and training purposes. The current study has focussed on the WSAP (as opposed to other measures of IB) as it is commonly used in training, whereas other measures focus on assessment (e.g., the Recognition Task; the Scrambled Sentences Task). The newly developed measure has been referred to as the Statement Word Association Task (SWAT). Participants were presented with a series of ambiguous statements and asked to resolve the ambiguity of each statement with a single word. Each item was reviewed by members of the research team, and we sought to validate the measure by rating the resolutions as either positive or negative in valence. The range of responses and the emotional valence generated were explored. We assessed internal consistency by examining split-half reliability. To assess convergent validity, the new measure of IB was evaluated against a number of existing measures related to IB,

anxiety, worry and depression. The measure was also given to participants at two different time-points to assess test-retest reliability.

1.4.1 Hypotheses

1. It was hypothesised that high levels of negative IB on the new worry-related SWAT measure would be associated with high levels of negative IB on an existing worry-related IB measure (criterion validity).
2. It was hypothesised that the higher the levels of negative IB on the new worry-related SWAT measure, the higher the reported anxiety, worry and depression (convergent validity).
3. It was hypothesised that there would be a unique contribution of worry in explaining the variance on the new measure of IB, whilst controlling for anxiety and depression (convergent validity).

2. Method

2.1 Participants

Participants were recruited from the crowdsourcing marketplace Amazon Mechanical Turk (MTurk) via the research platform CloudResearch (formerly TurkPrime). CloudResearch is an internet-based tool used in social sciences to collect behavioural data with speed and efficiency (Litman et al., 2017). It allowed the study to access additional recruitment features, including email communication with participants, additional control of study costs, and enhanced sampling options. The use of the MTurk platform also reduces the likelihood of responses generated either by bots (semi- or fully-automated code to respond automatically to surveys) or scripts (code that assists humans in responding more rapidly to certain questions; Kennedy et al. 2020). Research findings within health research have suggested such data quality issues are likely in studies in which participants are recruited through social media (Pozzar et al., 2020).

This study is interested in transdiagnostic cognitive processes, with worry conceptualised as an everyday common experience. CBM-I has been found to facilitate more positive interpretations and reduce negative intrusions in non-clinical populations (e.g., Hirsch et al., 2020). As such, people with varying levels of worry who may or may not receive a diagnosis of GAD were recruited from the general population. Information about participation and the nature of the research was accessible for prospective MTurk participants via a Qualtrics link. The inclusion criteria were people aged between 18-65, English as a native language, and English fluency. Participants were also required to be a resident in United States of America (USA), in order to avoid fraudulent responses that have emerged through MTurk in other countries (Kennedy et al., 2020). CloudResearch keeps track of participant histories and work quality through a number of indicators. As such, participants are granted an overall rating, which is updated every time they finish a study and is either rejected or approved by a

researcher/requester. This rating was used as selection criteria when specifying participant eligibility (i.e., inclusion of participants based on previous participation). Only participants with an approval rate ≥95% in their previous tasks on the platform were allowed to take part, to ensure good data quality. Participants were also required to have had ≥100 tasks previously approved. Exclusion criteria was English as a second language to a point whereby understanding of language may be limited.

The final sample consisted of N =146 participants at time point one (TP1) and N =134 for time point two (TP2). The sample size was akin to previous validation studies (e.g., Ogniewicz et al. 2014; Boffa et al. 2018; Rohrbacher & Reinecke, 2014). All participants reported English as their first language and USA as their country of residency.

2.2 Design

The study was a psychometric evaluation of a new measure using a correlational design with two time points, at least 2 weeks apart. The relationship between the new and an existing measure of IB was assessed. To assess criterion and convergent validity, the total score from the new measure of IB was correlated with measures of worry, anxiety, depression, and an existing IB measure. Hierarchical multiple regression was used to assess what proportion of the variance in the scores on the new measure were explained by worry, anxiety, and depression.

To assess reliability, internal consistency was examined by computing split-half correlations. Test-retest reliability was also assessed by computing intra-class correlation coefficients (ICC).

2.3 Materials and measures

2.3.1 Demographics questionnaire. This questionnaire included questions regarding participants' age, sex, highest level of education, occupation, language, and country of residency.

2.3.2 Worry measure. The Penn State Worry Questionnaire (PSWQ) was used to assess trait worry levels (Meyer et al., 1990). The participant is asked to rate 16 items on a scale from 1 (not at all typical of me) to 5 (very typical of me). Higher scores on the scale (after reverse-scoring appropriate items) denote greater worry. The PSWQ has good psychometric properties in student, community, and clinical samples, with studies reporting high internal consistency, short-term retest reliability, and convergent and criterion-related validity (Brown et al., 1992; Davey, 1993; Liu et al., 2021; Salarifar & Pouretemad, 2012). Cronbach's alphas in the present study were $\alpha = .97$ for TP1 and $\alpha = .90$ for TP2, denoting excellent internal consistency.

2.3.3 Depression symptoms. The Patient Health Questionnaire (PHQ) depression scale was used to measure symptoms of depression in the past 2 weeks. The nine-item PHQ-9 consists of the criteria on which the diagnosis of depressive disorders is based (American Psychiatric Association, 2013; Kroenke & Spitzer, 2002). Items are rated on a 4-point Likert scale ranging from 0 ("not at all") to 3 ("nearly every day"). Higher scores denote greater depression severity, and responses are summed to give an overall score. Evidence supports the reliability and validity of the PHQ-9 as a measure of depression in the general population (Kocalevent et al., 2013). Item 9 of the scale queries thoughts of death and self-harm, sometimes used as an indicator of suicide risk (Simon et al., 2016; Wu et al., 2020). For the purpose of the current study, the suicidality item was deemed not required and thus removed, and an adapted eight-item PHQ-8 used. Scores on the PHQ-8 and PHQ-9 have been found to be similar, with the PHQ-8 increasingly being used in research (Wu et al., 2020). Cronbach's

alphas in the present study were α = .87 for TP1 and α = .88 for TP2, denoting good internal consistency.

2.3.4 Anxiety symptoms. The Generalised Anxiety Disorder Assessment (GAD-7) was used to measure anxiety symptoms in the past 2 weeks. The seven-item scale consists of the criteria on which the diagnosis of generalised anxiety disorder is based (American Psychiatric Association, 2013; Spitzer et al., 2006). Items are rated on a 4-point Likert scale ranging from 0 ("not at all") to 3 ("nearly every day"). Responses are summed to give an overall score; higher scores denote greater anxiety severity. Evidence supports the reliability and validity of the GAD-7 as a measure of anxiety in the general population (Löwe et al., 2008). Cronbach's alphas in the present study were α = .89 for TP1 and α = .91 for TP2, denoting good and excellent internal consistency.

2.3.5 Existing measure of interpretation bias for worry. The Scrambled Sentences Task (SST) was adapted from Wenzlaff and Bates (1998) and involves reordering a list of five to six words to form grammatically correct sentences. The sentences can be either positive or negative in valence, and participants are asked to produce the first sentence that comes to mind. An example sentence includes "can my others merits faults see", which could be unscrambled to form the sentence "others can see my merits" (positive interpretation) or "others can see my faults" (negative interpretation). Participants were presented with 20 sentences to unscramble in 5 minutes, whilst holding a string of 6 digits in mind in order to suppress biased responding (Wenzlaff & Bates, 1998). The sentences selected relate to worry (Krahé et al., 2016; Krahé et al., under review). The outcome variable is the number of negative sentences divided by the total number of grammatically correct sentences generated. This serves as an index of IB, with a higher index (scores range from 0 to 1) denoting a more negative IB (Hirsch et al., 2018; Krahé et al., 2016; Krahé et al., 2019). Papers using the SST typically focus on split-half reliability to assess internal consistency (O'Connor et al., 2021; Würtz et al., 2021). Cronbach's

α may not be suitable to report, as the item coding is binary (see section 3.3.3 for split-half reliability coefficients).

2.3.6 Development of a new version of a measure of interpretation bias for worry. An IB statement resolution task for worry was developed, referred to as the Statement Word Association Task (SWAT). This measure differs from the SST, as it lends itself well to training IB (see Beard & Amir, 2009 and Hayes et al. 2010). The SWAT is an adaptation of the Word Sentence Association Paradigm (WSAP; Beard & Amir, 2009) mentioned in the Introduction (see section 1.3).

2.3.6.1 WSAP background information. In the WSAP, participants decide whether a word (implying a threat or benign interpretation) is related (or not related) to an ambiguous sentence. For example, a threat interpretation (e.g., "clumsy") or a benign interpretation (e.g., "graceful") precede an ambiguous sentence (e.g., "You carry a tray of food at a party"). The influence of positive and negative beliefs on interpreting ambiguous information is then examined (Beard & Amir, 2009). The original WSAP was developed for social anxiety (Beard & Amir, 2009), then for dysphoria (Sears et al., 2011), as well as for GAD and pathological worry (Ogniewicz et al., 2014). The developed GAD material was previously reviewed by Feng (2020) for consideration of potential inclusion for worry-related ambiguous statements. However, issues about the wide range of concepts (discussed later in this section) that can be activated with the ambiguous scenarios used in Ogniewicz et al. (2014)'s research were deemed problematic; that is, there was a lack of detail in the sentences. Indeed, the scenarios could be developed this way for the WSAP, as the target words were presented first, and can prime the interpretation before the scenario (Feng, 2020). This means the concept has been narrowed down by the word before the scenario, consequently making it easier to connect target words and the scenarios (even if they have broad concepts). The original WSAP was designed to take into account the tendency of individuals with anxiety disorders to rely on pre-existing negative

beliefs to resolve ambiguous information (Gonsalves et al., 2019). Studies that have adapted the WSAP and have presented resolutions opted to present the sentence first (Feng et al., 2019; Sears et al., 2011). This is because it seems more ecologically valid to encounter an ambiguous situation first, then generate a meaning of that situation, as per the new measure.

2.3.6.2 Completing the new SWAT measure. To complete the new SWAT measure, participants are presented with 120 ambiguous statements and asked to generate a one-word qualitative response for each statement, which they think resolves the ambiguity. An example includes, "You make a few errors in the driving test. At the end, the expression on the assessor's face indicates the result.". A positive response to this statement could be "passed" and a negative response could be "failed". Instructions for completing the SWAT were provided to participants (see appendix C), which were repeated halfway through alongside advice to take a short break. The order of the statements presented to participants was randomised and each statement was presented independently. Unlike the WSAP, there were no resolutions provided to participants and the statements were instead provided in isolation. This is crucial in the development of the task, as the words most frequently generated can then be used within training paradigms (i.e., a positive resolution to 'train' a more benign IB).

2.3.6.3 Development of the statements for the SWAT. The statements included in the SWAT were adapted from Feng et al. (2019) 's study. Feng et al. (2019) provided the research team with 120 existing ambiguous statements and 120 incomplete ambiguous statements, both related to worry. The incomplete statements require a single word at the end of the sentence to resolve the ambiguity; for example, "You have been asked to look after your niece because she is_____". For the purpose of developing the SWAT items, each scenario was initially reviewed (and either adapted or discarded) by the doctoral researcher. This then followed review by all members of the research team, including a member of the Liverpool Expert by Experience (LExE) team. This process was repeated 5 times, until 120 SWAT

statements were finalised (see appendix D). The LExE member within the research team then completed a pilot of the measure (within the entire online survey for both time points), at which point no further alterations were required.

At least 40 items were discarded from Feng et al. (2019)'s existing ambiguous statements list and replaced with an amended incomplete ambiguous statement. For example, an existing incomplete statement included "Your performance in the play was commented on by the other actors. What they said was very_____". This was replaced with "Your performance in the play was commented on by the other actors". In terms of existing ambiguous statements being discarded, reasons included there being no clear ambiguity in question (e.g., "You run to the train station and hope you do not miss the train. When you get there, you see the train"). Other reasons for items being discarded included the statement being unrelated to worry, appearing too "abstract" or "convoluted", or veering towards a particular valence. The remaining 80 existing ambiguous statements were also carefully reviewed. For example, statements were paraphrased when it was noted that an item did not "flow" easily when reading (e.g., "…you expect what it will be like to see them again" was amended to "…you think about what it will be like to see them again"). Other items were amended to ensure ambiguity (e.g., "You complete a test and as you look around the room, you see you are the last to finish" was changed to "You complete a test and as you look around the room, you wonder what it means if you are the last to finish"). It was also noted that some wording was specific to a British audience and was changed accordingly (e.g., "tap" was replaced with "faucet"; "council tax" was replaced with "property tax").

The process that informed the review and development of the SWAT items was as follows. Hirsch et al. (2016) suggest material for assessing IB should be developed close to the central component of the disorders, with different disorders having different concern topics (Hughes et al., 2016). Thus, content-specificity to the characteristics of worry was ensured,

with the ambiguous scenarios in the task relating to common worry domains (Feng, 2020). The worry domains, an example item, and the number of statements represented by each domain can be found in table 1.

Table 1. Worry domain, example item, and number of statements represented in the SWAT.

Worry domain	**Example item**	**Number of statements in the SWAT**
Relationship (social, family, or romantic)	*"You take some items back to the shop. The assistant agrees to give you a refund and you notice her attitude."*	25
Performance (work, academic, or others)	*"You are preparing for a test as part of a work diploma. You think about how you are likely to do."*	37
Physical harm or illness (self or others)	*"You hear footsteps running towards you. As they get closer, you turn around and see someone."*	13
Finance	*"Looking at your bank balance at the end of the month, you can see how you have managed your finances."*	17
Aimless future	*"As you reflect on your career goals and how much progress you have made so far, you realise how you are doing."*	12
Lack of confidence	*"You take your friend to a restaurant and recommend some of your favourite dishes to him. You know what he thinks of the taste by the expression on his face when he starts eating."*	7
General	*"You can see the bus at the bus stop and you are a little way off. You start to run and wave at the driver, and you know what the bus will do."*	9
TOTAL		120

Note. SWAT = Statement Word Association Task.

As per Feng et al. (2019)'s study, a number of rules were followed in the development of the SWAT items, in order to ensure face validity (see table 2). For example, IB may only occur when the situation is relevant to the person (rather than the viewpoint of someone else;

Hirsch et al. 2016). The statements were therefore developed to encourage self-referential processing, whereby the scenario describes realistic common situations that may occur to most people, or most people have experienced similar situations.

The developed scenarios allowed for both positive and negative solutions, and in each valence (positive or negative), participants should think of a similar word. It was therefore ensured that the individual items of the measure were constructed so that they funnel towards one specific concept. That is, if a scenario is not specific enough, a wide range of concepts may be activated by different people when encountering the ambiguous material. This means that even if the interpretation is of the same valence, it may not be of the same concept, and various ambiguities (rather than only one) may exist within the statement. Thus, participants might not be resolving the ambiguity in question. Alongside this, the interpretation is represented by the word in the context of the statement, rather than the meaning of the word in isolation (or a different ambiguity). For example, "You think about the number of life goals you are likely to achieve next year". A positive response to this statement could be "many" and a negative response could be "few", yet the valence would be unknown without knowledge of the statement. For the purpose of this thesis, the number of concepts generated was not explored in detail, therefore there were no statements discounted based on specificity of responses. However, statements were discarded if there was a high proportion of valence one-sidedness (e.g., mostly either positive or negative responses; see section 3.3.1 for item difficulty). If an item has mostly the same valence in its responses, this would suggest that the ambiguity is not open enough to both interpretations within the specific concept.

Table 2. Rules followed in the development of the SWAT (Feng et al., 2019).

Development of the SWAT items	
a)	The scenario is within the seven worry-related domains.
b)	The scenario describes realistic common situations that may occur to most people, or most people have experienced similar situations.
c)	The scenario allows both positive and negative solutions, and both solutions should have similar possibilities to be generated.
d)	The ambiguity of the scenario can be resolved by a single word, and in each valence (i.e., positive or negative) most people should think of a similar construct.
e)	The scenario is described in a self-referential way.
f)	The scenario is 50 words or shorter.

Note. SWAT = Statement Word Association Task.

2.3.6.4 Scoring of the SWAT. To mirror the SST, positive responses on the SWAT were given a score of 0 and negative responses are given a score of 1, with un-codable responses treated as missing data. The outcome measure involves the generated responses to be coded as either “negative”, “positive”, or “un-codable”. Table 3 details a decision log of the reasons items on the SWAT were deemed un-codable, including an example un-codable response alongside the corresponding statement.

Table 3. Decision log for participant responses deemed un-codable.

Decision Log for Un-codable Items	Example item	Example Un-codable response
Word seemingly not related to the statement and does not resolve ambiguity	*"You notice your partner's tone as they comment on how you have handled a difficult situation recently."*	*"Ready"*
Word related to the statement but without resolving the ambiguity	*"You overhear your manager saying something about you to another colleague."*	*"Eavesdropping"*
Word related to the statement but valence uncertain	*"You participate in an art competition. When you see the other entrant's work, you think about your own and know the chances of you winning the prize."*	*"Medium"*
Repetition of a word from the statement without resolving the ambiguity	*"You complete a test and as you look around the room, you wonder what it means if you are the last to finish."*	*"Wonder"*
More than a single word provided and without resolving the ambiguity	*"You've just started a new job and try to be useful. Today your boss wants to talk about your work productivity, and you know what she thinks."*	*"New tasks"*

2.3.6.5 Outcome variable of the SWAT. As for the SST, the outcome variable of the measure was the number of negative responses divided by the total number of codable responses. This serves as an index of IB, with a higher index (scores range from 0 to 1) denoting a more negative worry-related IB.

2.4 Procedure

Participants were invited to take part in a research study exploring the impact of mood and thinking style on performance on a sentence and word-based task. They were recruited from the general population using the US-based crowdsourcing platform MTurk. After logging

onto the MTurk site and selecting the session, prospective participants were made aware of inclusion and exclusion information. They were advised that the study involves x2 ca. 1-hour surveys to be completed at two separate time points, at least 2 weeks apart. A further breakdown of the nature of the tasks was provided, and a maximum time limit of 90 minutes for each time point was set. Participants were informed that they would be compensated $5.80 (US Dollars) for completion of each survey ($11.40 in total). Detail on ensuring payment was provided, whereby participants were required to share a unique code given at the end of the survey.

Participants who met criteria and decided to take part accessed the study using a Qualtrics link. Further information about the study was provided, including withdrawal procedures and information about the end of the study (e.g., dissemination and archiving; see appendix E for Participant Information Sheet). After providing informed consent (see appendix F), participants completed a short demographic questionnaire (see appendix G). The measures detailed in section 2.3 then followed. This firstly included the measure of worry, then anxiety and depression, followed by the existing and then the developed measure of IB under investigation. Participants who completed TP1 of the study were contacted via email 2 weeks later, in order to invite them to take part in TP2. This followed the same process as TP1, with the addition of debrief information (see appendix H).

The study asked questions related to worry, depression and anxiety, which can be sensitive areas. Support information was therefore provided at the end of the surveys (see appendix I).

2.5 Research ethics

The study protocol was approved by the Research Ethics Committee at the University of Liverpool (see appendix J). Following ethical approval, the study was pre-registered with

Open Science Framework (OSF), which can be accessed via the web address https://osf.io/m8j9a/. It should be noted that the final study deviated slightly from the original proposal on OSF. This includes the measure of depression as convergent (rather than discriminant), as well as no longer including measurement invariance within the analyses. The discussion section details the decision process with regards to the changes (see section 4.1 and 4.3 respectively).

2.6 Expert by experience consultation

A member of Liverpool Experts by Experience (LExE) provided consultation on all the materials used. This included review and feedback of the developed stimulus for the IB task (as detailed in section 2.3.6). Feedback was also provided on all participant information, and changes included ensuring clarity with regards to communicating withdrawal procedures. For example, participants were made aware that any data provided could be withdrawn up until the point of data analysis (at which point data was fully anonymised). LExE consultation also ensured that the information provided was accessible and inclusive. For example, amending certain words that could be considered complex (e.g., “ubiquitous”). A trial of both Qualtrics links for TP1 and TP2 also ensured that the surveys ran smoothly.

2.7 Data collection

Participants were recruited from July 2021 to August 2021. There was an interval of at least 2-weeks (and no greater than 3-weeks) between participants completing the measures for the first and second time.

2.8 Data analysis procedure

All participant responses on the SWAT were scored accordingly (i.e., positive responses = 0; negative responses = 1; un-codable responses treated as missing data). This meant that 17,520 responses were coded for TP1 (N = 146) and 16,080 responses were coded for TP2 (N = 134). Ten percent of the responses were coded by an independent rater, which included the first 12 items on the measure at both TP1 and TP2 (i.e., 3,384 responses coded in total). Intra-class correlation coefficients (ICC) were computed, and a high degree of inter-rater reliability was found for the SWAT between the raters. The average measure ICC was .99 with a 95% confidence interval from .99 to .99 (F (141, 141) = 362.813, p <.000).

Four participants were excluded from analysis from TP1. This was due to a high proportion (i.e., ≥57.5%) of un-codable items on the new measure of IB, as well as two of these participants completing ≤1 items on the SST.

Four participants were excluded from analysis from TP2. Two of the excluded participants provided a high proportion (i.e., 44.2% and 98%) of un-codable items on the new measure of IB, as well as completed ≤4 items on the SST. Another participant was excluded due to the repetition of only three words on the new measure of IB (i.e., “better”, “worse” or “affordable”), in addition to completing 0 items on the SST. A final participant was excluded due to providing 29.1% of un-codable items on the new measure of IB, as well as frequent repetition of responses (e.g., “good” or “nice”). This participant also completed only 4 items on the SST.

The participants excluded from TP1 either did not start (and therefore did not complete) TP2 or were otherwise also excluded at TP2 (for the reasons detailed above). If there were no other indicators for exclusion, participants who provided un-codable responses of ≤30% were still included for analysis.

2.9 Statistical analysis

Statistical analyses were carried out using SPSS version 27 and Stata 16 software (StataCorp, 2019). Data normality of the variables was assessed by consulting skewness and kurtosis values (Abbott, 2016).

The relationship between the new (SWAT) and existing (SST) measures of IB was assessed alongside the other variables, using Spearman's rank-order correlation coefficients (two-tailed). A scatterplot was also produced to display the strength of the relationship between the SWAT and the SST.

Hierarchical multiple regression was used to assess what proportion of the variance in the scores on the new measure were explained by worry, anxiety, and depression. To control for the contribution of depression and anxiety, the first step in the regression consisted of the PHQ-8 and GAD-7. As the study was interested in the contribution of worry in explaining the variance of the SWAT, the PSWQ was added at step 2.

Split-half reliability was calculated using Guttman's lambda (λ), in order to assess internal consistency. Guttman's λ considers all possible splits to find the maximal λ. A random series of locally optimal λ4 coefficients (1000 replications) was used and the following command was entered into Stata, "guttmanl var1-var100, pw quantiles(.05 .5 .95, rseed(1) reps(1000))". As per the command and to avoid overestimating reliability, the .5 quantile (median) alongside the maximal (1.0) λ4 value was reported.

In addition, intraclass correlation coefficients (ICC) were computed to assess the reliability of the new IB measure over time (i.e., differences between the mean SWAT index scores at TP1 and TP2). A two-way mixed effects model was performed with absolute agreement definition (i.e., column variance is excluded from denominator variance).

Table 4 details the type of reliability and validity considered within the study and the associated analysis computed.

Table 4. Type of reliability/validity and associated analysis.

Type of reliability/validity	Type of analysis
Convergent validity	Correlations (worry, anxiety, depression, existing IB measure)
Convergent validity	Hierarchical multiple regression (worry, anxiety, depression)
Split-half reliability	Split-half correlations
Test-retest reliability	Intra-class correlation coefficients (ICC)

3. Results

3.1 Demographic information

N = 142 participants were included in the analysis for TP1 and *N* = 130 participants for TP2. Mean age of the participants was *M* = 41.80 (SD = 10.48) for TP1 and *M* = 42.51 (*SD* = 10.45) for TP2. See table 5 for further demographic data. This includes a breakdown of gender ratios, as well as participant highest level of education.

Table 5. Participant demographic information.

Participant Demographics			
		TP1	TP2
Sample		*N* = 142	*N* = 130
Age (Mean)		41.80 (10.48)	42.51 (10.45)
Female		70	64
Male		72	66
Highest level of education	Bachelor's degree	57.7%	59.2%
	High school graduate	24.6%	23.8%
	Master's degree	11.3%	10.8%
	Associate's degree	4.9%	4.6%
	Other (e.g., vocational training)	1.4%	1.5%

Note. Standard deviations (SD) in parenthesis. TP1= time point 1; TP2 = time point 2.

There was an equal proportion of male and females that completed the surveys at both time points, and the most commonly reported highest level of education was Bachelor's degree.

3.2 Descriptive statistics

Table 6 shows participant means, SDs, range, and median scores for each measure, at both TP1 and TP2. Additional SWAT index scores have been presented due to 20 statements

being omitted from the measure following item difficulty (see section 3.3.1); however, the scores remained consistent.

Table 6. Descriptive statistics for measures.

Descriptive Statistics for Measures						
	Means (SD)		Range		Median	
	TP1	TP2	TP1	TP2	TP1	TP2
SWAT index	.39 (.20)	.37 (.21)	.90	.96	.40	.34
SWAT index †	.37 (.19)	.39 (.22)	.84	.96	.38	.38
SST index	.29 (.25)	.24 (.24)	1.00	1.00	.20	.20
PSWQ total	46.39 (17.32)	45.70 (13.31)	64.00	48.00	44.50	44.00
GAD-7 total	4.20 (4.61)	3.73 (4.60)	19.00	18.00	3.00	2.00
PHQ-8 total	3.84 (4.48)	3.61 (4.45)	19.00	19.00	3.00	2.00

Note. Standard deviations (SD) in parenthesis. TP1= time point 1; TP2 = time point 2. SWAT = Statement Word Association Task; SST = Scrambled Sentences Task; PSWQ = Penn State Worry Questionnaire; GAD-7 = Generalised Anxiety Disorder Assessment; PHQ-8 = Patient Health Questionnaire depression scale.

† *20/120 statements omitted following item difficulty.*

Mean scores for anxiety and depression were low, and typical of what would be expected within the general population. That is, a cut-off point of ≥10 on the PHQ-8 can be used for defining current depression (Kroenke et al., 2009), and cut-off scores of 7-10 on the GAD-7 for identifying GAD (Plummer et al., 2016). The range of scores on the PSWQ was large (possible range of scores is 16-80) and reduced from 64 at TP1 to 48 at TP2. The mean scores for worry at both time points were indicative of moderate worriers (i.e., scores of 16-39 low worry; 40-59 moderate worry; 60-80 high worry; Meyer et al., 1990).

The measures of IB produced an index score ranging from 0 to 1, with lower scores indicating a more negative IB. The mean scores on the SST were relatively low, indicating a more benign IB, and slightly reduced from .29 at TP1 to .24 at TP2. The mean index scores on the SWAT were also low but somewhat higher than mean index scores on the SST. Indeed,

future studies are needed to establish clinical cut-offs of measures of IB (Gonsalves et al., 2019); although it would be expected that most people within the general population would have a fairly benign IB.

3.3 Further development of the SWAT

3.3.1 Item difficulty. In order to examine the proportion of valence one-sidedness (i.e., the majority of responses coded as either positive or negative), a frequency table of the individual items of the SWAT was generated. This is because the items should allow for both positive and negative responses, and the validity of the item would be questioned should there be a high proportion of responses in either direction. The table also examined the frequency of un-codable responses per SWAT item, as a high number of un-codable responses would question item clarity. Percentages were calculated for the entire dataset at both time points for "positive", "negative", or "un-codable" responses. A full list of the items can be found in appendix K. Table 7 shows an example of the percentages calculated for the first 1-5 items of the dataset.

Table 7. Frequency plot of the first five items of the SWAT.

Item number	Positive IB		Negative IB		Un-codable	
	TP1	TP2	TP1	TP2	TP1	TP2
1	43.7%	53.8%	52.1%	43.8%	4.2%	2.3%
2	52.1%	64.6	44.4%	31.5%	3.5%	3.8%
3	40.1%	37.7%	54.2%	54.6%	5.6%	7.7%
~~4~~	~~18.3%~~	~~18.5%~~	~~78.2%~~	~~80.0%~~	~~3.5%~~	~~1.5%~~
5	71.5%	71.5%	23.9%	26.9%	4.9%	1.5%

Note. IB = interpretation bias; TP1 = time point 2; TP2 = time point 2.

Item difficulty occurred when a low proportion of responses were found to be either negative or positive (cut-off chosen at ≥20%). This meant that participants were likely to

provide a response with the same valence, whereas what was aimed for was a balance of both positive and negative responses. As such, 20 statements were omitted from the SWAT, as indicated by a strikethrough the item on the table. An example of an omitted statement includes item 4, where on average only 18.4% of participants responded with a positive word. The ambiguous statement for item 4 was, "You are completing a group project, but one of the team members is not pulling their weight. You decide to talk about this, and as you start to speak, you know what this conversation will be like". It is possible that this statement led the participants to respond negatively, rather than allowing for ambiguity in both directions. This statement could be amended, such as, "You are completing a group project and decide to talk about group productivity. As you start to speak, you know what this conversation will be like". This might then limit item suggestibility and increase the ambiguity in question.

Item difficulty also occurred when a high proportion of responses were found to be un-codable (cut-off chosen at ≤20%); however, there were no items omitted on this basis (including within the items already omitted due to item one-sidedness). This suggests that all items included an ambiguity that could be resolved by a single word.

To explore item consistency across the two time points, the positive IB and negative IB percentages per item were reviewed. The positive IB percentages and the negative IB percentages for each item were subtracted, to examine the difference in scores between the two time points. The mean percentage difference for positive IB was M = 3.1% (SD = 5.1%) and the mean percentage difference for negative IB was M = 1.8% (SD = 4.7%). The percentage frequencies suggest a high degree of consistency across time points in terms of the emotional valence generated.

A total of 100 out of originally 120 statements were taken forward and included for further analyses.

3.3.2 Individual item response range. To give examples of the range of responses that were generated on an item, two illustrative word clouds are shown in figures 1 and 2. The larger the word in size, the higher the response rate for that word. A variety of different words were found on each individual item, as demonstrated on both word clouds.

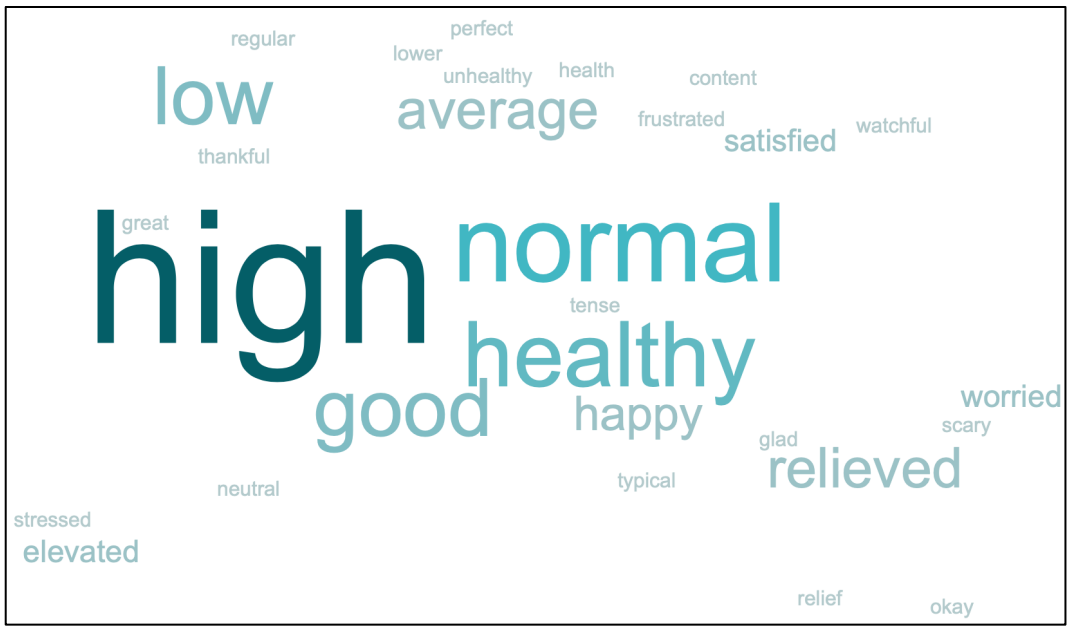


Figure 1. Illustrative word cloud for SWAT item 9 for time point 1.
Note. SWAT item 9: "The value on the blood pressure meter indicates your blood pressure".

Figure 2. Illustrative word cloud for SWAT item 112 for time point 1.
Note. SWAT item 112: "You are looking for a birthday present for your mother. You see one that your mother will like, and you check the price tag".

As can be seen in the first example, participants mostly generated “high” (negative interpretation), followed by “normal” (positive interpretation). In the second example, participants mostly generated “affordable” (positive interpretation), followed by “expensive” (negative interpretation). This means the items worked as intended; that is, participants chose similar words within a different emotional valence. Corresponding frequency graphs of the words that were generated on the word clouds can be found in figures 3 and 4.

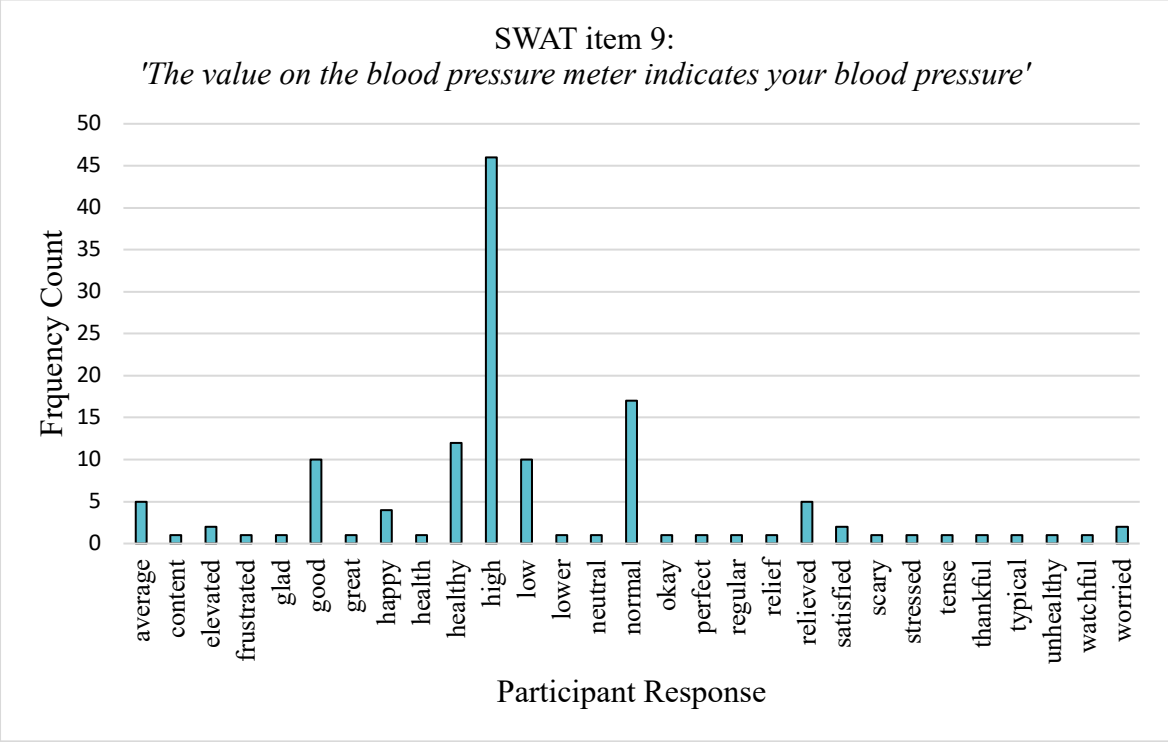


Figure 3. Frequency graph for SWAT item 9 for time point 1.
Note. N = 133. SWAT = Statement Word Association Task. Participant responses presented A-Z.

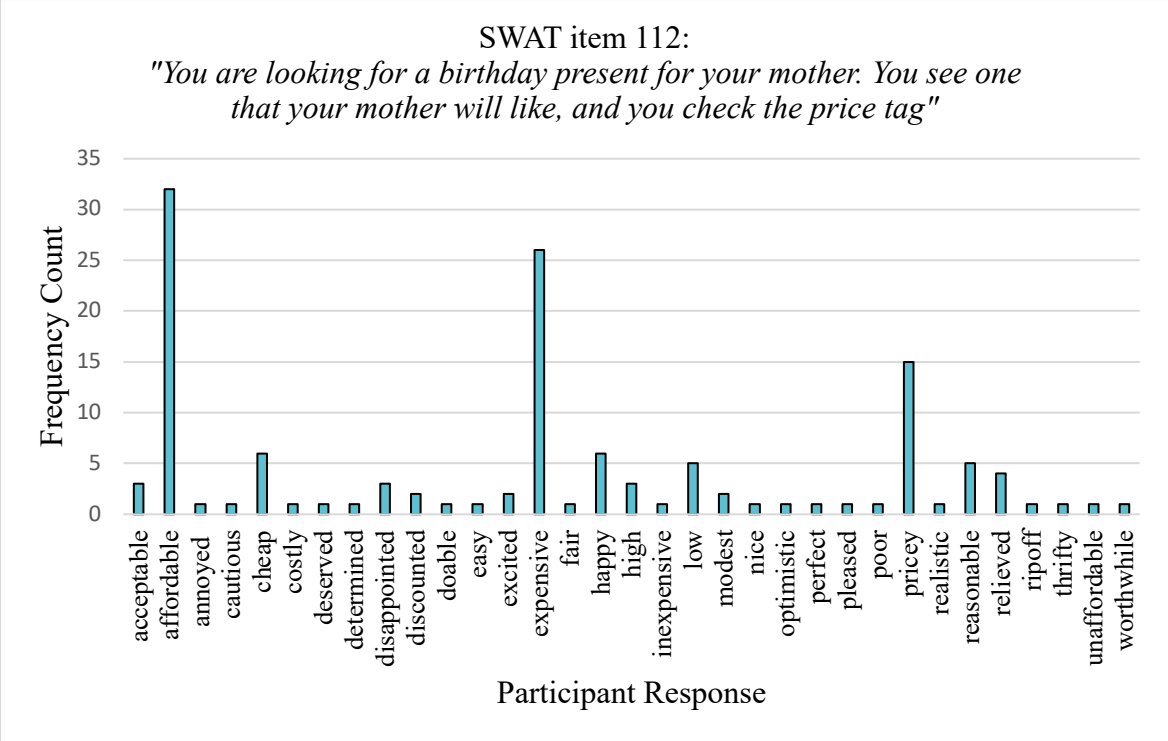


Figure 4. Frequency graph for SWAT item 112 for time point 1.
Note. N = 133. SWAT = Word Sentence Association Task. Participant responses presented A-Z.

The graphs demonstrate how often participants responded with the same words. In the first example, “high” (negative interpretation), “normal” (positive interpretation) and “low” (positive interpretation) were the most frequent responses. In the second example, “affordable” (positive interpretation), “expensive” (negative interpretation) and “pricey” (negative interpretation) were the most frequent responses. Although the word with the highest response count differs in valence, the graphs show more variety in the positive responses for both items. This suggests that the negative concept is potentially better defined within these statements.

All analyses were repeated for TP2 and corresponding word clouds and frequency graphs can be found in appendix L. The results at TP2 were comparable to TP1.

3.3.3 Split-half reliability of the SWAT and SST. Cronbach’s α is the most popular measure of reliability reported in psychological research, yet often underestimates test reliability (Dunn et al., 2014). The most common approach to compute Cronbach’s α involves averaging the correlations between each item’s score and the sum score of the remaining items.

However, this assumes the trial stimuli are presented in a fixed order, thus not a valid estimate of reliability for the SWAT (Parsons et al., 2019), as the items on the SWAT were presented in a random order (i.e., no two participants received the same order of items). The SWAT also produced 'missing' data, which can lead to further difficulties when computing internal consistency (e.g., methods using casewise deletion).

There have been numerous alternative indices proposed as being superior to Cronbach's α for test reliability. Estimating measurement reliability with a permutation-based approach has been advised, whereby the data are repeatedly randomly split into two halves and the reliability estimate calculated for each split (Parsons et al., 2019). Guttman's lambda (λ) was therefore calculated as a measure of split-half reliability, which considers all possible splits to find the maximal λ. A random series of locally optimal λ4 coefficients (1000 replications) was used. To avoid overestimating reliability, the .5 quantile (median) alongside the maximal (1.0) λ4 value was reported. Pairwise rather than casewise was used to deal with missing values. Guttman's λ4 was excellent for the SWAT at .99 (.5 quantile = .98) for TP1 and .99 (.5 quantile = .98) for TP2. For the SST, guttman's λ4 was excellent at .87 (.5 quantile = .83) for TP1 and .89 (.5 quantile = .84) for TP2.

3.4 The relationship between the SWAT and the measures of interpretation bias, worry, anxiety and depression

3.4.1 Convergence among variables. The relationship between the measures were investigated using correlation coefficients. Preliminary analyses were initially performed to examine any violation of the assumptions of normality, linearity, and homoscedasticity. Most of the measures were skewed at both time points (aside from the SWAT); thus, Spearman's rank-order correlational analysis was used. As per individual item screen, 100 items from the SWAT were included for all inferential statistical analyses.

As hypothesised, a moderate positive linear association (r = .67, n = 142, p <.001; see table 8) was found between the two measures of IB at TP1, as shown in figure 5. As expected, there was a strong, positive correlation found between the SWAT and SST (r = .67, n = 142, p <.001), with high levels of negative IB on the SST associated with high levels of IB as measured by the SWAT.

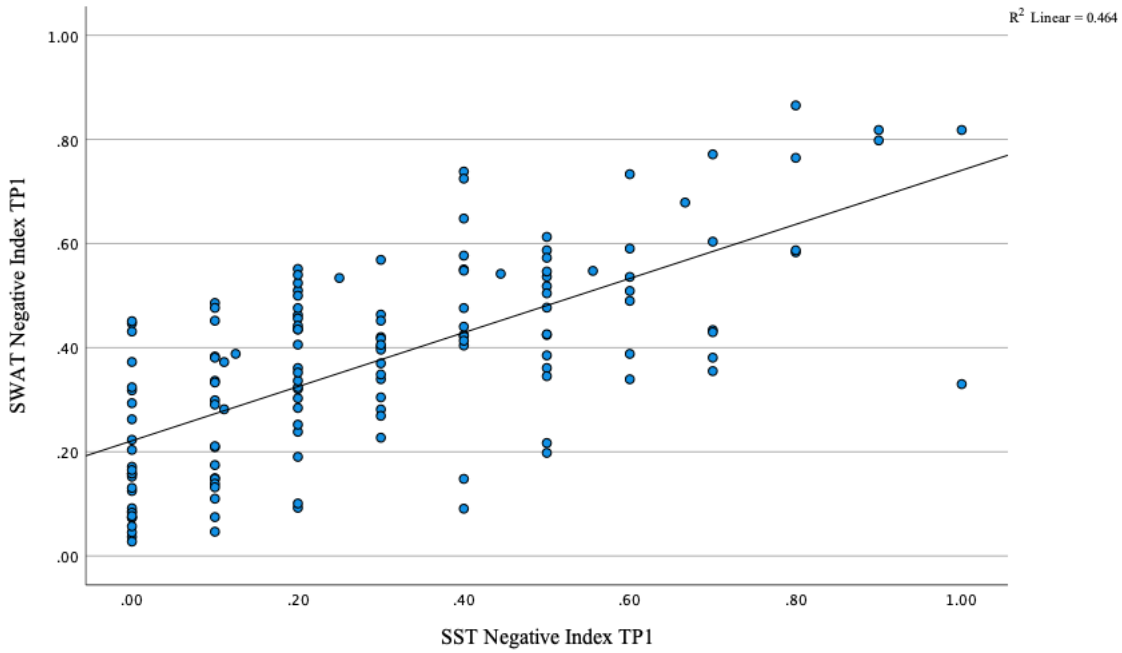


Figure 5. Association between the SWAT index scores and SST index scores for time point 1.
Note. N = 142. SWAT = Statement Word Association Task; SST = Scrambled Sentences Task.

As hypothesised, significant correlations were found between all the measures at TP1, as presented in table 8. The correlations between the SWAT and the PSWQ (r = .49, n = 142, p <.001), GAD-7 (r = .48, n = 142, p <.001), and PHQ-8 (r = .45, n = 142, p <.001) were all moderate. Coefficient of determination showed that negative IB on the SST helps to explain nearly 46% of the variance of negative IB on the SWAT.

Table 8. *Correlation matrix of the main variables for time point 1.*

Spearman's Rho Correlations Between Measures					
Measures	SWAT	SST	PSWQ	GAD-7	PHQ-8
SWAT					
SST	.667*				
PSWQ	.494*	.592*			
GAD-7	.478*	.596*	.754*		
PHQ-8	.449*	.573*	.762*	.762*	

Note. N = 142. SWAT = Statement Word Association Task; SST = Scrambled Sentences Task; PSWQ = Penn State Worry Questionnaire; GAD-7 = Generalised Anxiety Disorder Assessment; PHQ-8 = Patient Health Questionnaire depression scale.
** = correlation is significant at p <.001*

3.4.2 Assessing predictors of the SWAT. A hierarchical multiple regression was conducted with the SWAT as the outcome variable. This was in order to control for the contribution of depression and anxiety, as the study was interested in the unique contribution of worry in explaining the variance of the SWAT. There was no violation of independence of errors (Durban-Watson = 1.720) and no concerns about multicollinearity (VIF range = 1.952 to 3.086).

The first stage of the regression consisted of the PHQ-8 and GAD-7. This regression model accounted for a significant proportion of the variance, F (2, 139) = 21.197, p <.001, R^2 = .234. The second stage consisted of adding the PSWQ. This model accounted for a significant proportion of the variance, F (1, 138) = 13.197, p <.001, R^2 = .301. The model as a whole (i.e., the contribution of anxiety, depression and worry) explained 30.1% of the variance in the SWAT.

The PSWQ explained an additional 6.7% of the variance in the SWAT, even when the effects of the PHQ-8 and GAD-7 were statistically controlled for. This change in R^2 was significant, F(3, 138) = 19.770, p <.001. The results of the individual predictors are presented

in table 9. When all three predictors were included in stage 2 of the regression model, GAD-7 was no longer a statistically significant predictor of the SWAT.

The PSWQ made the strongest unique contribution to explaining the scores on the SWAT, when the variance explained by all other variables in the model is controlled for. There was no in priori assumption of how demographic variables should be related to the outcome and thus not controlled for in the hierarchical regression.

Table 9. Results from a hierarchical regression assessing predictors of the SWAT for time point 1.

Hierarchical regression with the SWAT as the Criterion Variable

Source	Unstandardised β coefficients	Standard error	Standardised β coefficients	*t*	Sig.	R^2	Adj. R^2	R^2 change
Step 1						.234	.223	.234
Constant	.286	.020		14.631	<.001			
PHQ-8	0.10	.004	.228	2.196	.030			
GAD-7	0.012	.004	.296	2.853	.005			
Step 2						.301	.285	.067
Constant	.140	.044		3.156	.002			
PHQ-8	.009	.004	.216	2.170	.032			
GAD-7	.001	.005	.021	.166	.869			
PSWQ	.004	.001	.384	3.633	<.001			

Note. N = 142. SWAT = Statement Word Association Task; PSWQ = Penn State Worry Questionnaire; GAD-7 = Generalised Anxiety Disorder Assessment; PHQ-8 = Patient Health Questionnaire depression scale.

Simple linear regressions were carried out to determine influence of demographic information on SWAT scores. There was no statistically significant model found for age (F (1, 140) = .027, p = .869, R^2 = .00), education (F (1, 140) = .166, p = .684, R^2 = .001), or sex (F (1, 140) = .188, p = .665, R^2 = .001).

All analyses were repeated for TP2 and the corresponding correlation and hierarchical multiple regression tables can be found in appendix M, as well as the SWAT and SST

scatterplot. The results were comparable to TP1, aside from the hierarchical multiple regression analyses. That is, when all three predictors were included in stage 2 of the regression model at TP2, both GAD-7 and the PHQ-8 (rather than just GAD-7) were no longer a statistically significant predictor of the SWAT. This would suggest that worry is a better predictor of the SWAT, as hypothesised, especially at TP2.

3.5 The performance of the SWAT over time

3.5.1 Test-retest reliability of the SWAT. Intra-class correlation coefficients (ICC) were computed to assess the reliability of the SWAT over time.

A high degree of reliability was found for the SWAT between TP1 and TP2. The average measure ICC was .94 with a 95% confidence interval from .91 to .96 (F (129, 129) = 16.672, p <.000), indicating excellent test-retest reliability.

4. Discussion

There is evidence within the literature demonstrating that individuals with high levels of worry tend to make threatening interpretations when appraising ambiguous information (Feng et al., 2019; Hirsch & Mathews, 2012; Krahé et al., 2019). Yet, there is a paucity of research regarding the reliability and validity of the measures used. The current study therefore aimed to adapt and validate a new version of the WSAP, as a measure of worry-related IB that can be used for both assessment and training purposes. Referred to as the SWAT, the measure yielded positive results with regards to its psychometric properties. Included within the study are findings related to content validity, criterion validity, convergent validity, test-retest reliability, inter-rater reliability, and internal consistency. An overview and discussion of the findings, clinical implications, limitations, and future directions are detailed.

4.1 Overview and discussion of findings

Content validity of the SWAT was considered within the development stage of the measure. It refers to the degree to which the elements of the new measure are representative and relate to worry (i.e., the targeted construct), including clarity of instructions, adequacy of the response format, and linguistic aspects of the items (Haynes et al., 1995; Koller et al., 2017). Content validity is determined via expert judgement. As such, the research team were familiar with the central components and characteristics of worry, as informed by previous research (Dugas et al., 1998; Hirsch et al., 2018; Tallis et al., 1992). The characteristics of worry were reviewed and represented by seven common worry domains (Feng, 2020). Alongside this, the 'rules' for the development of the items clearly set out what was required of each item, to ensure face validity (e.g., content-specificity; self-referential processing). There was also an awareness of neutrality within the statement formation development and review; that is, ambiguity that seeks to not be leading towards the emotional valence of the resolution. The

research team provided expert opinion and judgment in the area of IB, including extensive clinical and research experience of working with people with high levels of worry. Future research could seek to pilot the measure or conduct a focus group, in order to evaluate whether participants felt the statements were representative of their worries.

Descriptive statistics further ensured the measure allowed for the resolution of ambiguity in both directions (i.e., positive and negative). This included the omission of items whereby valence preference emerged (and prior to any inferential statistical analyses). Unlike previous IB measures, the nature of the SWAT did not provide any choice for respondents (e.g., providing an option of words for deciding what would best resolve the ambiguity). This meant that instructions for completion of the measure were slightly more comprehensive than previous measures, as examined during the development phase. The response rate of un-codable items was low, suggesting clarity in the instructions and individual items generated. Participants were therefore likely to have understood what was required for completion of the measure and thus practical usage was ensured. A run through of the SWAT was completed by all members of the research team, including a member of LExE. This not only ensured that the Qualtrics links ran smoothly, but also for the purpose of further ensuring face validity.

The review of valence one-sidedness in the initial stages of analyses allowed for 100 items to remain on the SWAT, suggesting suitability for the measurement of IB in worry. Convergent validity was subsequently demonstrated by the patterns of correlations between the remaining SWAT items and the existing measures. High levels of negative IB on the SWAT were associated with high levels of negative IB on the SST, as per hypothesis 1 and supporting criterion validity. Significant correlations were also found between the SWAT and the existing measures of anxiety, worry and depression, as per hypothesis 2. This suggested that the new measure relates to other measures of the same and/or similar constructs.

The key finding of the current study was that trait worry was found to be closely related to the SWAT. However, it should be noted that an amendment to the original proposal involved the inclusion of depression as convergent, rather than discriminant, as noted in the OSF pre-registration. This is because further review of the literature indicated a clear overlap of the constructs. That is, worry and rumination are conceptualised as part of a transdiagnostic repetitive-negative-thinking process (Ehring & Watkins, 2008); and rumination is a common characteristic of depression (Nolen-Hoeksema et al., 2008). A recent study by O'Connor et al. (2021) examined the convergence among IB measures and their associations with depressive symptom severity. As important to the current study, the SST bias index was found to be uniquely related to depression and potentially the best marker of depression-linked IB (O'Connor et al., 2021). Regarding hypothesis 3, hierarchical regression analysis showed that worry made the strongest unique contribution to explaining the scores on the SWAT. This provides evidence for the association of worry and IB, which few studies to date have demonstrated (e.g., Krahé et al., 2019 and Feng et al., 2019). It was interesting that depression remained statistically significant when all three predictors were included, whilst anxiety did not. This could have implications with regards to theoretical models of anxiety and provide a rationale for separating out the cognitive features. That is, worry is understood to be a defining feature of GAD; yet the results indicate something unique between the two constructs.

4.2 Clinical implications

Interpretation modification techniques are considered a promising target when developing effective interventions for people experiencing emotional psychopathology (Beard, 2011; Hirsch et al., 2018). Indeed, cognitive therapeutic approaches for anxiety often aim to reduce threatening ways of interpreting situations (Beck, 1976; Beck & Clark, 1997; Mathews & Mackintosh, 1998). Experimental research has since shown that inducing negative

interpretations in healthy individuals elevates anxiety, whilst inducing positive interpretations reduces anxiety (Mackintosh et al., 2006; Mathews & Mackintosh, 2000; Mathews & MacLeod, 2002). This has led to the development of computerised training of more positive interpretations styles, as a form of intervention for anxiety (i.e., CMB-I; Ji et al. (2021)).

CBM-I has been found to be helpful for people with a diagnosis of GAD or depression, when focused on worry or rumination materials, respectively (Hirsch et al., 2018). Research has also found that internet-based CBM-I can reduce longer-term repetitive negative thinking (Hirsch et al., 2020). Yet, there remain issues with the methodologies in CBM-I research, whereby the measures people used are often not properly validated. It is therefore hoped that the SWAT will be a useful addition to the field, providing an appropriate measure of IB for worry, to be used for assessment and training purposes. As described by Gonsalves et al. (2019), it could be employed similarly to existing WSAP measures when using the tool as a CBM-I task. That is, the measure is unique whereby it could present both benign and threat interpretations (see section 4.4 for details on the resolutions provided by participants being used in future studies). This means that individuals could be given the opportunity to notice from the items on the measure that there are multiple ways to interpret the same situation (Beard et al., 2019). It is possible that practicing the skill of rejecting interpretations may be particularly helpful (Gonsalves et al., 2019). This contrasts with the ambiguous scenario training task that is frequently used (e.g., Mathews and Mackintosh, 2000), which only present positive resolutions to ambiguous scenarios. Indeed, both types of CBM-I have demonstrated efficacy (Gonsalves et al., 2019). It is posited that internet-delivered IB training could offer a low cost, accessible, and non-labour intensive intervention for targeting IB and anxiety symptom severity (Ji et al., 2021). Hirsch et al. (2020) suggests that future research could determine whether CBM-I could augment CBT outcomes.

4.3 Limitations

The sample size of the current study was informed by previous studies, alongside budget limitations in the context of participant payment costs. However, it is acknowledged that there are other validation studies with a higher number of participants (e.g., Orchard et al. (2016); Gonzalez et al. (2017)). Indeed, there appears limited guidance within the literature in terms of the sample size required for validation studies. A review by Anthoine et al. (2014) found that the sample size determination for psychometric validation studies is rarely ever justified a *priori*; thus, emphasising a lack of clear scientifically sound recommendations within the area. Approximately 92% of the articles within Anthoine et al. (2014)'s review reported a subject to item ratio greater than or equal to 2. This ratio is much lower in the current study; however, analyses that might have required more data (e.g., a confirmatory factor analysis) were not undertaken. Recruitment via MTurk meant that there was relative ease with regards to participant interest and involvement, whereby there were limited issues with recruitment. Although concerns could be raised in terms of heterogeneity within such online recruitment, MTurk participants (referred to as 'workers' on the platform) typically represent a diverse group (Stritch et al., 2017). The ease of access to MTurk samples thus allows research on the validation of single measures and multiple constructs (Stritch et al., 2017). Nevertheless, the generalisability of the results from the current study could be questioned due to the high proportion of degree-level participants that took part. That is, the SWAT measure requires participants to resolve the ambiguity of a grammatically complete statement, which might require a certain level of cognitive ability.

The original proposal of the current study included measurement invariance in order to test the structural validity (i.e., psychometric equivalence) of the SWAT across time (Putnick & Bornstein, 2016). It was considered that such analyses would be fundamental towards ensuring the measure could be used as an assessment for pre/post training purposes (e.g., within

experimental research). As well, the metrics of structural validity has been demonstrated as severely underreported in psychological research, suggesting widespread hidden invalidity of measures used (Hussey & Hughes, 2020). However, the process of reporting measurement invariance was further reviewed at the analyses stage of the study. It was noted that assumptions cannot be made with regards to there being just one underlying latent factor for the number of items developed for the SWAT. Indeed, measurement invariance requires specification of the number of latent factors in the measure, which is unknown for the SWAT without further investigation (i.e., exploratory factor analysis). Whilst it is likely that a general worry factor would be found, the measure was designed to tap into lots of different dimensions of worry, meaning other factors could emerge (e.g., finance; aimless future). A confirmatory factor analysis would therefore not be possible. The outcome variable for the SWAT is also binary and measurement invariance would typically involve metric scales. It was therefore deemed not appropriate to include measurement invariance, or indeed integral for validation of the measure, particularly when considering future utility. Should the measure be used as a training tool, it is less important to establish how many underlying latent factors should emerge, as the SWAT is looking to target a whole range of different worriers. Indeed, a high degree of reliability of the SWAT over time was already indicated by test-retest reliability. Excellent internal consistency at each time point was also reported.

A final limitation to consider is the nature of the scoring required for the SWAT. That is, it requires the rater to use their own judgement in terms of coding the responses (i.e., positive, negative, or un-codable). This could mean that different coding could be found amongst different people rating the responses; however, inter-rater reliability in the current study suggested excellent agreement among coding. It is therefore suggested that the decision log for participant responses deemed un-codable will offer guidance for future users of the

measure. It may also be helpful to consult additional opinion if there is uncertainty on the individual scoring of an item.

4.4 Future directions

It was interesting that the two individual item graphs that were explored in the current study showed more variety in positive responses in terms of response range. Whilst this could suggest that the negative concept is potentially better defined within these statements, the difference could also be understood within the density hypothesis (see Unkelbach et al. 2008). The core assumption of this hypothesis is that positive information is more alike other positive information, compared to the overall similarity of negative information (Unkelbach, 2012). This higher similarity can then be translated to higher density in spatial representations of information. That is, the words "happy" and "pleased" are more similar than "sad" and "horrible"; consequently, the positive words are more densely clustered (i.e., located closer together) than the negative words, if plotted in graphical representations (Unkelbach, 2012). Most languages are more diversified on the negative side (Schrauf & Sanchez, 2004); yet, there may be fewer similar words that could negatively resolve an ambiguity within the same construct. This could be an alternative explanation in terms of increased variety of positive responses. On the other hand, it should be acknowledged that the responses on the SWAT are unique in terms of considering valence dimension; whereby they can only be made sense of in conjunction with the statement.

The word clouds and frequency graphs provided a snapshot into the range of words that were generated per item. The current study grouped (and scored) these words in terms of emotional valence, with the view that the words would likely be of a similar construct. For example, the responses "acceptable", "affordable", "fair" and "inexpensive" from item 112 could reasonably be seen to fall within a similar category. However, the data was not explored

in terms of the concepts generated per item, beyond that of the valence. Although outside the scope and indeed purpose of this thesis, the data generated could allow for a rich qualitative exploration of the different conceptual categories generated.

In terms of utility, it is also hoped that the SWAT can be used in future studies, including as a measure of IB for worry whereby people are asked to make a relatedness judgement (i.e., supraliminally). However, limitations exist with regards to understanding the initiation and maintenance of worry in GAD; whereby it does not distinguish whether IB occurs on reflection, or indeed when ambiguity is first encountered (i.e., an automatic response; Hirsch et al. (2016)). Participants still see ambiguous material and resolutions in current online paradigms (i.e., they are consciously aware of the resolution, even if the point is not to rank resolutions). Thus, a key area for future direction concerning worry-related IB involves experimental research. For example, it is known that subliminally presenting information can have an influence on participants (Gibbons, 2009). Subliminal priming occurs when an individual is exposed to stimuli below the threshold of perceptions (Barutchu et al., 2018; Kawakami et al., 2018). This is a process that occurs outside the realm of consciousness (Elgendi et al., 2018). Indeed, research using electroencephalography (EEG) recording has shown that worry-related interpretations seem to occur at an early stage of information processing (Feng et al., 2019); and it is possible that interpretation modification could also occur at this stage. Understanding and predicting the psychological states (e.g., worry) of participants in response to subliminal priming could inform our understanding of the processes by which interpretations are made and their time course. The SWAT items could be used as a measure that tests the possibility that interpretations can be trained outside of conscious awareness (i.e., subliminally). That is, a future study that seeks to tease apart at what point when encountering ambiguity an interpretation is made (i.e., a proof-of-concept study).

The words that were generated within the current study could be incorporated within

the subliminal training paradigm, offering a crowd-developed feature of the measure. That is, participants who are higher worriers are subliminally presented positive words/resolutions to the SWAT items, in order to see whether a training effect to promote a more positive bias occurs. A threshold finding task to identify the subliminal threshold for each participant would be performed at the start of the training paradigm. The word would then be presented first, whereby a subliminal cue could modify the development of the interpretation as the sentence is read. The positively trained group would be compared to a control condition that involves subliminally presenting neutral resolutions unrelated to the ambiguity (i.e., 'sham trained'). The developed SWAT measure included in such research would allow further understanding into the stages of processing at which people make interpretations. This could then have implications for further research into potential novel treatments for GAD (Hirsch et al., 2018). Indeed, consideration would be required with regards subliminal processing and the words presented (e.g., word length; syllable count; frequency of use within the English language).

5. Conclusions

The current study provides support for the use of the SWAT as a measure to be used in future research on worry and IB. It also provides evidence for the association between IB and worry, as per the results from both the new and existing IB measures. Adaptations of the developed measure involved the omission of items that did not allow for both the positive and negative resolution of the statements. Recommendations for a future subliminal training paradigm was discussed, whereby the validation of the SWAT offers a pre-cursor to the rigour of such experimental research. Future studies should continue to ensure reliability and validity of the methods used within IB research. This includes detailed methodology with regards to the development of the measures.

6. References

Abbott, M. L. (2016). *Using statistics in the social and health sciences with SPSS and excel*. John Wiley & Sons.

Ambady, N., & Gray, H. M. (2002). On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology, 83*(4), 947.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). American Psychiatric Association.

Anderson, K. G., Dugas, M. J., Koerner, N., Radomsky, A. S., Savard, P., & Turcotte, J. (2012). Interpretive style and intolerance of uncertainty in individuals with anxiety disorders: A focus on generalized anxiety disorder. *Journal of Anxiety Disorders, 26*(8), 823-832.

Anthoine, E., Moret, L., Regnault, A., Sébille, V., & Hardouin, J.-B. (2014). Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes, 12*(1), 1-10.

Barker, C., Pistrang, N., & Elliott, R. (2016). Foundations in quantitative measurement. In C. Barker, N. Pistrang, & R. Elliott (Eds.), *Research methods in clinical psychology: An introduction for students and practitioners*. John Wiley & Sons.

Barutchu, A., Spence, C., & Humphreys, G. W. (2018). Multisensory enhancement elicited by unconscious visual stimuli. *Experimental Brain Research, 236*(2), 409-417.

Beard, C. (2011). Cognitive bias modification for anxiety: current evidence and future directions. *Expert Review of Neurotherapeutics, 11*(2), 299-311.

Beard, C., & Amir, N. (2009). Interpretation in social anxiety: When meaning precedes ambiguity. *Cognitive Therapy and Research, 33*(4), 406-415.

Beard, C., Rifkin, L. S., Silverman, A. L., & Björgvinsson, T. (2019). Translating CBM-I into real-world settings: Augmenting a CBT-based psychiatric hospital program. *Behavior Therapy, 50*(3), 515-530.

Beck, A. T. (1976). Cognitive therapy and the emotional disorders (Vol. 8). In *Cognitive therapy and the emotional disorders*. Penguin.

Beck, A. T., & Clark, D. A. (1997). An information processing model of anxiety: Automatic and strategic processes. *Behaviour Research and Therapy, 35*(1), 49-58.

Boffa, J. W., Norr, A. M., Tock, J. L., Amir, N., & Schmidt, N. B. (2018). Development of the interpretation bias index for PTSD. *Cognitive Therapy and Research, 42*(5), 720-734.

Borkovec, T. D., Alcaine, O., & Behar, E. (2004). Avoidance theory of worry and generalized anxiety disorder. In R. G. Heimberg, C. L. Turk, & D. S. Mennin (Eds.), Generalized anxiety disorder: advances in research and practice. Guilford Press.

Borkovec, T., Newman, M. G., Pincus, A. L., & Lytle, R. (2002). A component analysis of cognitive-behavioral therapy for generalized anxiety disorder and the role of interpersonal problems. *Journal of Consulting and Clinical Psychology, 70*(2), 288.

Borkovec, T. D., Robinson, E., Pruzinsky, T., & DePree, J. A. (1983). Preliminary exploration of worry: Some characteristics and processes. *Behaviour Research and Therapy, 21*(1), 9-16.

Brown, T. A., Antony, M. M., & Barlow, D. H. (1992). Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. *Behaviour Research and Therapy, 30*(1), 33-37.

Butler, G., & Mathews, A. (1983). Cognitive processes in anxiety. *Advances in Behaviour Research and Therapy, 5*(1), 51-62.

Colman, M. A. (2006). *Oxford Dictionary of Psychology* (2nd ed.). Oxford University Press.

Davey, G. C. (1993). A comparison of three worry questionnaires. *Behaviour Research and Therapy, 31*(1), 51-56.

Davey, G. C., & Meeten, F. (2016). The perseverative worry bout: A review of cognitive, affective and motivational factors that contribute to worry perseveration. *Biological Psychology, 121*, 233-243.

Dillon, K. H., Allan, N. P., Cougle, J. R., & Fincham, F. D. (2016). Measuring hostile interpretation bias: The WSAP-hostility scale. *Assessment, 23*(6), 707-719.

Dugas, M. J., Freeston, M. H., Ladouceur, R., Rhéaume, J., Provencher, M., & Boisvert, J.-M. (1998). Worry themes in primary GAD, secondary GAD, and other anxiety disorders. *Journal of Anxiety Disorders, 12*(3), 253-261.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399-412.

Ehring, T., & Watkins, E. R. (2008). Repetitive negative thinking as a transdiagnostic process. *International Journal of Cognitive Therapy, 1*(3), 192-205.

Elgendi, M., Kumar, P., Barbic, S., Howard, N., Abbott, D., & Cichocki, A. (2018). Subliminal priming—state of the art and future perspectives. *Behavioral Sciences, 8*(6), 54.

Everaert, J., Podina, I. R., & Koster, E. H. (2017). A comprehensive meta-analysis of interpretation biases in depression. *Clinical Psychology Review, 58*, 33-48.

Eysenck, M. W., Mogg, K., May, J., Richards, A., & Mathews, A. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of Abnormal Psychology, 100*(2), 144.

Feng, Y.-C. (2020). *Understanding the cognitive mechanisms of worry and interpretation training using behavioural and Event-Related Potentials (ERPs) markers* (Doctoral Dissertation, King's College London).

Feng, Y.-C., Krahé, C., Koster, E. H. W., Lau, J. Y., & Hirsch, C. R. (under review). Cognitive processes predict worry and anxiety under different stressful situations.

Feng, Y.-C., Krahé, C., Sumich, A., Meeten, F., Lau, J. Y., & Hirsch, C. R. (2019). Using event-related potential and behavioural evidence to understand interpretation bias in relation to worry. *Biological Psychology, 148*, 107746.

Gibbons, H. (2009). Evaluative priming from subliminal emotional words: Insights from event-related potentials and individual differences related to anxiety. *Consciousness and Cognition, 18*(2), 383-400.

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.

Gonzalez, A., Rozenman, M., Langley, A. K., Kendall, P. C., Ginsburg, G. S., Compton, S., Walkup, J. T., Birmaher, B., Albano, A. M., & Piacentini, J. (2017). Social interpretation bias in children and adolescents with anxiety disorders: psychometric examination of the self-report of ambiguous social situations for youth (SASSY) scale. *Child & Youth Care Forum, 46*(3), 395-412.

Grey, S., & Mathews, A. (2000). Effects of training on interpretation of emotional ambiguity. *The Quarterly Journal of Experimental Psychology: Section A, 53*(4), 1143-1162.

Hayes, S., Hirsch, C. R., Krebs, G., & Mathews, A. (2010). The effects of modifying interpretation bias on worry in generalized anxiety disorder. *Behaviour Research and Therapy, 48*(3), 171-178.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247.

Hazlett-Stevens, H., & Borkovec, T. (2004). Interpretive cues and ambiguity in generalized anxiety disorder. *Behaviour Research and Therapy, 42*(8), 881-892.

Hertel, P. T., & Mathews, A. (2011). Cognitive bias modification: Past perspectives, current findings, and future applications. *Perspectives on Psychological Science, 6*(6), 521-536.

Hirsch, C. R., Hayes, S., & Mathews, A. (2009). Looking on the bright side: accessing benign meanings reduces worry. *Journal of Abnormal Psychology, 118*(1), 44.

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.

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.

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.

Hirsch, C. R., & Mathews, A. (2000). Impaired positive inferential bias in social phobia. *Journal of Abnormal Psychology, 109*(4), 705.

Hirsch, C. R., & Mathews, A. (2012). A cognitive model of pathological worry. *Behaviour Research and Therapy, 50*(10), 636-646.

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.

Hughes, A. M., Gordon, R., Chalder, T., Hirsch, C. R., & Moss-Morris, R. (2016). Maximizing potential impact of experimental research into cognitive processes in health psychology: A systematic approach to material development. *British Journal of Health Psychology, 21*(4), 764-780.

Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science, 3*(2), 166-184.

In-Albon, T., Klein, A., Rinck, M., Becker, E., & Schneider, S. (2008). Development and evaluation of a new paradigm for the assessment of anxiety-disorder-specific interpretation bias using picture stimuli. *Cognition and Emotion, 22*(3), 422-436.

Ji, J. L., Baee, S., Zhang, D., Calicho-Mamani, C. P., Meyer, M. J., Funk, D., Portnow, S., Barnes, L., & Teachman, B. A. (2021). Multi-session online interpretation bias training for anxiety in a community sample. *Behaviour Research and Therapy*, 103864.

Kawakami, N., Miura, E., & Nagai, M. (2018). When you become a superman: Subliminal exposure to death-related stimuli enhances men's physical force. *Frontiers in Psychology, 9*, 221.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P., Jewell, R., & Nicholas, W. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods*(8), 614-629.

Kertz, S. J., & Woodruff-Borden, J. (2011). The developmental psychopathology of worry. *Clinical Child and Family Psychology Review, 14*(2), 174-197.

Kocalevent, R.-D., Hinz, A., & Brähler, E. (2013). Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. *General Hospital Psychiatry, 35*(5), 551-555.

Koller, I., Levenson, M. R., & Glück, J. (2017). What do you think you are measuring? A mixed-methods procedure for assessing the content validity of test items and theory-based scaling. *Frontiers in Psychology, 8*, 126.

Koster, E. H., Fox, E., & MacLeod, C. (2009). Introduction to the special section on cognitive bias modification in emotional disorders. *Journal of Abnormal Psychology, 118*(1), 1.

Krahé, C., Mathews, A., Whyte, J., & Hirsch, C. R. (2016). Cognitive bias modification for interpretation with and without prior repetitive negative thinking to reduce worry and rumination in generalised anxiety disorder and depression: protocol for a multisession experimental study with an active control condition. *British Medical Journal Open, 6*(12), e013404.

Krahé, C., Meeten, F., & Hirsch, C. R. (under review). Development and psychometric evaluation of a scrambled sentences test specifically for worry in individuals with generalized anxiety disorder.

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.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric Annals, 32*(9), 509-515.

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders, 114*(1-3), 163-173.

Ladouceur, R., Talbot, F., & Dugas, M. J. (1997). Behavioral expressions of intolerance of uncertainty in worry: Experimental findings. *Behavior Modification*, *21*(3), 355-371.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433-442.

Liu, K., Nijmeh, J. S., & Warren, S. L. (2021). Factor Structure, Measurement Invariance, and Concurrent Validity of the Penn State Worry Questionnaire Across Development, Psychopathology, and Culture. *Assessment*, 1-16.

Llera, S. J., & Newman, M. G. (2010). Effects of worry on physiological and subjective reactivity to emotional stimuli in generalized anxiety disorder and nonanxious control participants. *Emotion, 10*(5), 640.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, 266-274.

Mackintosh, B., Mathews, A., Yiend, J., Ridgeway, V., & Cook, E. (2006). Induced biases in emotional interpretation influence stress vulnerability and endure despite changes in context. *Behavior Therapy, 37*(3), 209-222.

Martinelli, M. K., Holzinger, J. B., & Chasson, G. S. (2014). Validation of an interpretation bias assessment for body dissatisfaction. *Body Image, 11*(4), 557-561.

Mathews, A., & Mackintosh, B. (1998). A cognitive model of selective processing in anxiety. *Cognitive Therapy and Research, 22*(6), 539-560.

Mathews, A., & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of Abnormal Psychology, 109*(4), 602-615.

Mathews, A., & MacLeod, C. (2002). Induced processing biases have causal effects on anxiety. *Cognition & Emotion, 16*(3), 331-354.

Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology, 1*, 167-195.

Mathews, A., Richards, A., & Eysenck, M. (1989). Interpretation of homophones related to threat in anxiety states. *Journal of Abnormal Psychology, 98*(1), 31.

Mennin, D. S., Heimberg, R. G., Turk, C. L., & Fresco, D. M. (2005). Preliminary evidence for an emotion dysregulation model of generalized anxiety disorder. *Behaviour Research and Therapy*, *43*(10), 1281-1310.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy, 28*(6), 487-495.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to

complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100.

NICE. (2011). *Generalised anxiety disorder and panic disorder in adults: management*. National Institute for Health and Care Excellence Guideline (CG113).

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science, 3*(5), 400-424.

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*, 1-16.

Ogniewicz, A. S., Dugas, M. J., Langlois, F., Gosselin, P., & Koerner, N. (2014). An adapted word-sentence association paradigm for generalized anxiety and worry: assessing interpretation bias. *Journal of Experimental Psychopathology, 5*(4), 457-476.

Orchard, F., Pass, L., & Reynolds, S. (2016). Associations between interpretation bias and depression in adolescents. *Cognitive Therapy and Research, 40*(4), 577-583.

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.

Plummer, F., Manea, L., Trepel, D., & McMillan, D. (2016). Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. *General Hospital Psychiatry, 39*, 24-31.

Pozzar, R., Hammer, M. J., Underhill-Blazey, M., Wright, A. A., Tulsky, J. A., Hong, F., Gundersen, D. A., & Berry, D. L. (2020). Threats of bots and other bad actors to data quality following research participant recruitment through social media: Cross-sectional questionnaire. *Journal of Medical Internet research, 22*(10), e23021.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90.

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.

Ruscio, A. M. (2002). Delimiting the boundaries of generalized anxiety disorder: Differentiating high worriers with and without GAD. *Journal of Anxiety Disorders, 16*(4), 377-400.

Salarifar, M., & Pouretemad, H. (2012). P-1060-The study of factorial structure, validity and reliability of the penn state worry questionnaire (PSWQ). *European Psychiatry, 27*(S1), 1-1.

Schoth, D. E., & Liossi, C. (2017). A systematic review of experimental paradigms for exploring biased interpretation of ambiguous information with emotional and neutral associations. *Frontiers in Psychology, 8*, 171.

Schrauf, R. W., & Sanchez, J. (2004). The preponderance of negative emotion words in the emotion lexicon: A cross-generational and cross-linguistic study. *Journal of Multilingual and Multicultural Development, 25*(2-3), 266-284.

Sears, C. R., Bisson, M. S., & Nielsen, K. E. (2011). Dysphoria and the immediate interpretation of ambiguity: Evidence for a negative interpretive bias in error rates but not response latencies. *Cognitive Therapy and Research, 35*(5), 469-476.

Sibrava, N. J., & Borkovec, T. D. (2006). The cognitive avoidance theory of worry. In G. C. Davey & A. Wells (Eds.), *Worry and its psychological disorders: Theory, assessment and treatment* (pp. 239-258). John Wiley & Sons.

Simon, G. E., Coleman, K. J., Rossom, R. C., Beck, A., Oliver, M., Johnson, E., Whiteside, U., Operskalski, B., Penfold, R. B., & Shortreed, S. M. (2016). Risk of suicide attempt and suicide death following completion of the Patient Health Questionnaire depression module in community practice. *The Journal of Clinical Psychiatry, 77*(2), 221-227.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097.

StataCorp. (2019). *Stata Statistical Software: Release 16*. StataCorp LLC.

Stritch, J. M., Pedersen, M. J., & Taggart, G. (2017). The opportunities and limitations of using Mechanical Turk (Mturk) in public administration and management scholarship. *International Public Management Journal, 20*(3), 489-511.

Szabó, M. (2011). The emotional experience associated with worrying: anxiety, depression, or stress? *Anxiety, Stress, and Coping, 24*(1), 91-105.

Tallis, F., Eysenck, M., & Mathews, A. (1992). A questionnaire for the measurement of nonpathological worry. *Personality and Individual Differences, 13*(2), 161-168.

Unkelbach, C. (2012). Positivity advantages in social information processing. *Social and Personality Psychology Compass, 6*(1), 83-94.

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: the density hypothesis. *Journal of Personality and Social Psychology, 95*(1), 36-49.

Wells, A. (1995). Meta-cognition and worry: A cognitive model of generalized anxiety disorder. *Behavioural and Cognitive Psychotherapy, 23*(3), 301-320.

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.

Wu, Y., Levis, B., Riehm, K. E., Saadat, N., Levis, A. W., Azar, M., Rice, D. B., Boruff, J., Cuijpers, P., & Gilbody, S. (2020). Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis. *Psychological Medicine, 50*(8), 1368-1380.

Würtz, F., Zahler, L., Blackwell, S. E., Margraf, J., Woud, M., & Würtz, F. (2021). Scrambled but valid? The Scrambled Sentences Task as a Measure of Interpretation Biases in Psychopathology: A Systematic Review. *PsyArXiv*.

Xu, S., Ouyang, X., Shi, X., Li, Y., Chen, D., Lai, Y., & Fan, F. (2020). Emotional exhaustion and sleep-related worry as serial mediators between sleep disturbance and depressive symptoms in student nurses: A longitudinal analysis. *Journal of Psychosomatic Research, 129*, 109870.

Appendices

Appendix A. A measurement tool to assess systematic reviews (AMSTAR 2).

1. Did the research questions and inclusion criteria for the review include the components of PICO?

For Yes:
- ☐ Population
- ☐ Intervention
- ☐ Comparator group
- ☐ Outcome

Optional (recommended)
- ☐ Timeframe for follow-up

☐ Yes
☐ No

2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?

For Partial Yes:
The authors state that they had a written protocol or guide that included ALL the following:
- ☐ review question(s)
- ☐ a search strategy
- ☐ inclusion/exclusion criteria
- ☐ a risk of bias assessment

For Yes:
As for partial yes, plus the protocol should be registered and should also have specified:
- ☐ a meta-analysis/synthesis plan, if appropriate, *and*
- ☐ a plan for investigating causes of heterogeneity
- ☐ justification for any deviations from the protocol

☐ Yes
☐ Partial Yes
☐ No

3. Did the review authors explain their selection of the study designs for inclusion in the review?

For Yes, the review should satisfy ONE of the following:
- ☐ *Explanation for* including only RCTs
- ☐ OR *Explanation for* including only NRSI
- ☐ OR *Explanation for* including both RCTs and NRSI

☐ Yes
☐ No

4. Did the review authors use a comprehensive literature search strategy?

For Partial Yes (all the following):
- ☐ searched at least 2 databases (relevant to research question)
- ☐ provided key word and/or search strategy
- ☐ justified publication restrictions (e.g. language)

For Yes, should also have (all the following):
- ☐ searched the reference lists / bibliographies of included studies
- ☐ searched trial/study registries
- ☐ included/consulted content experts in the field
- ☐ where relevant, searched for grey literature
- ☐ conducted search within 24 months of completion of the review

☐ Yes
☐ Partial Yes
☐ No

5. Did the review authors perform study selection in duplicate?

For Yes, either ONE of the following:
- ☐ at least two reviewers independently agreed on selection of eligible studies and achieved consensus on which studies to include
- ☐ OR two reviewers selected a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder selected by one reviewer.

☐ Yes
☐ No

6. Did the review authors perform data extraction in duplicate?

For Yes, either ONE of the following:

- ☐ at least two reviewers achieved consensus on which data to extract from included studies
- ☐ OR two reviewers extracted data from a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder extracted by one reviewer.

☐ Yes
☐ No

7. Did the review authors provide a list of excluded studies and justify the exclusions?

For Partial Yes:

- ☐ provided a list of all potentially relevant studies that were read in full-text form but excluded from the review

For Yes, must also have:

- ☐ Justified the exclusion from the review of each potentially relevant study

☐ Yes
☐ Partial Yes
☐ No

8. Did the review authors describe the included studies in adequate detail?

For Partial Yes (ALL the following):

- ☐ described populations
- ☐ described interventions
- ☐ described comparators
- ☐ described outcomes
- ☐ described research designs

For Yes, should also have ALL the following:

- ☐ described population in detail
- ☐ described intervention in detail (including doses where relevant)
- ☐ described comparator in detail (including doses where relevant)
- ☐ described study's setting
- ☐ timeframe for follow-up

☐ Yes
☐ Partial Yes
☐ No

9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?

RCTs

For Partial Yes, must have assessed RoB from

- ☐ unconcealed allocation, *and*
- ☐ lack of blinding of patients and assessors when assessing outcomes (unnecessary for objective outcomes such as all-cause mortality)

For Yes, must also have assessed RoB from:

- ☐ allocation sequence that was not truly random, *and*
- ☐ selection of the reported result from among multiple measurements or analyses of a specified outcome

☐ Yes
☐ Partial Yes
☐ No
☐ Includes only NRSI

NRSI

For Partial Yes, must have assessed RoB:

- ☐ from confounding, *and*
- ☐ from selection bias

For Yes, must also have assessed RoB:

- ☐ methods used to ascertain exposures and outcomes, *and*
- ☐ selection of the reported result from among multiple measurements or analyses of a specified outcome

☐ Yes
☐ Partial Yes
☐ No
☐ Includes only RCTs

10. Did the review authors report on the sources of funding for the studies included in the review?

For Yes

- ☐ Must have reported on the sources of funding for individual studies included in the review. Note: Reporting that the reviewers looked for this information but it was not reported by study authors also qualifies

☐ Yes
☐ No

11. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?

RCTs

For Yes:

- ☐ The authors justified combining the data in a meta-analysis
  - ☐ AND they used an appropriate weighted technique to combine study results and adjusted for heterogeneity if present.
  - ☐ AND investigated the causes of any heterogeneity

☐ Yes
☐ No
☐ No meta-analysis conducted

For NRSI

For Yes:

- ☐ The authors justified combining the data in a meta-analysis
  - ☐ AND they used an appropriate weighted technique to combine study results, adjusting for heterogeneity if present
  - ☐ AND they statistically combined effect estimates from NRSI that were adjusted for confounding, rather than combining raw data, or justified combining raw data when adjusted effect estimates were not available
  - ☐ AND they reported separate summary estimates for RCTs and NRSI separately when both were included in the review

☐ Yes
☐ No
☐ No meta-analysis conducted

12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?

For Yes:

- ☐ included only low risk of bias RCTs
- ☐ OR, if the pooled estimate was based on RCTs and/or NRSI at variable RoB, the authors performed analyses to investigate possible impact of RoB on summary estimates of effect.

☐ Yes
☐ No
☐ No meta-analysis conducted

13. Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?

For Yes:

- ☐ included only low risk of bias RCTs
- ☐ OR, if RCTs with moderate or high RoB, or NRSI were included the review provided a discussion of the likely impact of RoB on the results

☐ Yes
☐ No

14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?

For Yes:

- ☐ There was no significant heterogeneity in the results
- ☐ OR if heterogeneity was present the authors performed an investigation of sources of any heterogeneity in the results and discussed the impact of this on the results of the review

☐ Yes
☐ No

15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?

For Yes:

- ☐ performed graphical or statistical tests for publication bias and discussed the likelihood and magnitude of impact of publication bias

☐ Yes
☐ No
☐ No meta-analysis conducted

16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?

For Yes:

- ☐ The authors reported no competing interests OR
- ☐ The authors described their funding sources and how they managed potential conflicts of interest

☐ Yes
☐ No

Appendix B. Full article reviewed and excluded records.

Unavailable in English

(1) Couturier, J., Pellegrini, D., Miller, C., Bhatnagar, N., Boachie, A., Bourret, K., Brouwers, M., Coelho, J. S., Dimitropoulos, G., Findlay, S., Ford, C., Geller, J., Grewal, S., Gusella, J., Isserlin, L., Jericho, M., Johnson, N., Katzman, D. K., Kimber, M., Lafrance, A., Leclerc, A., Loewen, R., Loewen, T., McVey, G., Norris, M., Pilon, D., Preskow, W., Spettigue, W., Steinegger, C., Waite, E., & Webb, C. (2021, Apr). The COVID-19 pandemic and eating disorders in children, adolescents, and emerging adults: virtual care recommendations from the Canadian consensus panel during COVID-19 and beyond. *Journal of Eating Disorders, 9*(1), Article 46.

(2) Hay, P. J., Claudino, A. M., Touyz, S., & Abd Elbaky, G. (2015, Jul 27). Individual psychological therapy in the outpatient treatment of adults with anorexia nervosa. *Cochrane Database of Systematic Reviews, 2015*(7), Cd003909.

(3) Hay, P. J., Sud, R., & Touyz, S. (2012, Dec 2012). Treatment for severe and enduring anorexia nervosa: A review. *Australian and New Zealand Journal of Psychiatry, 46*(12), 1136-1144.

University dissertation

(4) Hale, K., Ho, J., Cohen, A., & Sazgar, S. (2016). Neuropsychological rehabilitation specific to anorexia nervosa: a critical review of the literature on executive functioning symptomatology and cognitive remediation treatment applications tailored to this patient population. *Clinical Neuropsychologist, 30*(3), 422-423.

Search strategy not stated

(5) Attia, E. (2021). Anorexia nervosa treatment trials: time for new approaches. *The Lancet Psychiatry, 8*(3), 170-171.

(6) Baldock, E., & Tchanturia, K. (2007, May 2007). Translating laboratory research into practice: Foundations, functions and future of cognitive remediation therapy for anorexia nervosa. *Therapy, 4*(3), 285-292.

(7) Berg, K. C., & Wonderlich, S. A. (2013, Nov). Emerging psychological treatments in the field of eating disorders. *Current Psychiatry Reports, 15*(11), Article 407.

(8) Cardi, V., Tchanturia, K., & Treasure, J. (2018). Premorbid and illness-related social difficulties in eating disorders: An overview of the literature and treatment developments. *Current Neuropharmacology, 16*(8), 1122-1130.

(9) Cooper, Z., & Bailey-Straebler, S. (2015, Mar). Disseminating evidence-based psychological treatments for eating disorders. *Current Psychiatry Reports, 17*(3), Article 12.

(10) Couturier, J. (2018, Oct 2018). Can family-based treatment for eating disorders in youth be improved? *Journal of the American Academy of Child and Adolescent Psychiatry, 57*(10).

(11) Danner, U. N., Dingemans, A. E., & Steinglass, J. (2015, Nov). Cognitive remediation therapy for eating disorders. *Current Opinion in Psychiatry, 28*(6), 468-472.

(12) Eichen, D. M., Matheson, B. E., Appleton-Knapp, S. L., & Boutelle, K. N. (2017, Sep). Neurocognitive treatments for eating disorders and obesity. *Current Psychiatry Reports, 19*(9), Article 62.

(13) Herpertz-Dahlmann, B., van Elburg, A., Castro-Fornieles, J., & Schmidt, U. (2015). ESCAP Expert Paper: New developments in the diagnosis and treatment of adolescent anorexia nervosa—a European perspective [Review]. *European Child and Adolescent Psychiatry, 24*(10), 1153-1167.

(14) Juarascio, A. S., Manasse, S. M., Espel, H. M., Kerrigan, S. G., & Forman, E. M. (2015). Could training executive function improve treatment outcomes for eating disorders? *Appetite, 90*, 187-193.

(15) Kass, A. E., Kolko, R. P., & Wilfley, D. E. (2013, Nov). Psychological treatments for eating disorders. *Current Opinion in Psychiatry, 26*(6), 549-555.

(16) Kelly, N. R., Shank, L. M., Bakalar, J. L., & Tanofsky-Kraff, M. (2014). Pediatric feeding and eating disorders: Current state of diagnosis and treatment. *Current Psychiatry Reports, 16*(5), Article 446.

(17) Kim, E. J., Bahk, Y. C., Oh, H., Lee, W. H., Lee, J. S., & Choi, K. H. (2018, Oct). Current status of cognitive remediation for psychiatric disorders: A review. *Frontiers in Psychiatry, 9*, Article 461.

(18) La Buissonniere-Ariza, V., Schneider, S. C., & Storch, E. A. (2017). Cognitive remediation of executive functioning in youth with neuropsychiatric conditions: current knowledge on feasibility, effectiveness, and personalization. *Expert Review of Precision Medicine and Drug Development, 2*(2), 133-144.

(19) Martinez, M. A., & Craighead, L. W. (2015, Sep). Toward Person(ality)-centered treatment: How consideration of personality and individual differences in anorexia nervosa may improve treatment outcome. *Clinical Psychology-Science and Practice, 22*(3), 296-314.

(20) Mitchell, K., & Carr, A Carr, Alan (2000). Anorexia and bulimia. In A. Carr (Eds.). *What works with children and adolescents? A critical review of psychological interventions with children, adolescents and their families*. Routledge.

(21) Moynihan, J., Giombini, L., & Nesbitt, S. (2016, Jul). A systematic evaluation of cognitive remediation therapy (CRT) for the treatment of children and adolescents with anorexia nervosa. *European Eating Disorders Review, 24*(4), E8-E8.

(22) Murray, S. B. (2020). Updates in the treatment of eating disorders in 2019: a year in review in eating disorders: The journal of treatment & prevention. *Eating Disorders, 28*(1), 21-31.

(23) Plemeniti Tololeski, B., Suhodolcan Grabner, A., & Kumperscak, H. G. (2021). Adolescents with autism spectrum disorder and anorexia nervosa comorbidity: Common features and treatment Possibilities with cognitive remediation therapy and oxytocin. *Frontiers in Psychiatry, 12*.

(24) Tchanturia, K., & Hambrook, D. (2010). Cognitive remediation therapy for anorexia nervosa. *The treatment of eating disorders: A clinical handbook*, 130-735.

(25) Tchanturia, K., & Lock, J. (2010). Cognitive remediation therapy for eating disorders: development, refinement and future directions. *Behavioral Neurobiology of Eating Disorders*, 269-287.

(26) Wonderlich, S. A., Bulik, C. M., Schmidt, U., Steiger, H., & Hoek, H. W. (2020). Severe and enduring anorexia nervosa: Update and observations about the current clinical reality. *International Journal of Eating Disorders, 53*(8), 1303-1312.

(27) Zhu, J., Yang, Y., Touyz, S., Park, R., & Hay, P. (2020). Psychological treatments for people with severe and enduring anorexia nervosa: A mini review. *Front Psychiatry, 11*, 206.

(28) Zipfel, S., Giel, K. E., Bulik, C. M., Hay, P., & Schmidt, U. (2015). Anorexia nervosa: Aetiology, assessment, and treatment [Review]. *The Lancet Psychiatry, 2*(12), 1099-1111.

Focused on other treatment modalities

(29) Barnett, P., Arundell, L. L., Saunders, R., Matthews, H., & Pilling, S. (2021). The efficacy of psychological interventions for the prevention and treatment of mental health disorders in university students: A systematic review and meta-analysis [Review]. *Journal of Affective Disorders, 280*, 381-406.

(30) Churchill, R., Moore, T. H., Furukawa, T. A., Caldwell, D. M., Davies, P., Jones, H., Shinohara, K., Imai, H., Lewis, G., & Hunot, V. (2013). 'Third wave' cognitive and behavioural therapies versus treatment as usual for depression. *Cochrane Database of Systematic Reviews, 2013*(10), CD008705.

(31) Gregertsen, E. C., Mandy, W., Kanakam, N., Armstrong, S., & Serpell, L. (2019). Pre-treatment patient characteristics as predictors of drop-out and treatment outcome in individual and family therapy for adolescents and adults with anorexia nervosa: A systematic review and meta-analysis [Review]. *Psychiatry Research, 271*, 484-501.

(32) Hannah, L., Cross, M., Baily, H., Grimwade, K., Clarke, T., & Allan, S. M. A systematic review of the impact of carer interventions on outcomes for patients with eating disorders. *Eating and Weight Disorders-Studies on Anorexia Bulimia and Obesity, 16,* 1-10.

(33) Langmesser, L., & Verscheure, S. (2009). Are eating disorder prevention programs effective. *Journal of Athletic Training, 44*(3), 304-305.

(34) Lock, J. (2015, Sep). An update on evidence-based psychosocial treatments for eating disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 44*(5), 707-721.

(35) Mann, S. E., Erford, B. T., & Minnich, A. (2018). Meta-analysis of anorexia nervosa counseling outcome studies from 1990 to 2015. *Journal of Counseling and Development, 96*(1), 41-52.

(36) Nagata, J. M., & Murray, S. B. (2021). Updates in the treatment of eating disorders in 2020: a year in review in eating disorders: the journal of treatment & prevention [Article]. *Eating Disorders, 29*(2), 123-133.

(37) Reche, M., & Ivars, D. (2015). Anorexia nervosa: Treatment review. *European Psychiatry, 30*, 1355.

(38) Robertson, A., & Thornton, C. (2021). Challenging rigidity in anorexia (treatment, training and supervision): questioning manual adherence in the face of complexity. *Journal of Eating Disorders, 9*(1), Article 104.

(39) Solmi, M., Favaro, A., Wade, T. D., Johnson, C., Byrne, S., Del Giovane, C., Fairburn, C. G., Ostinelli, E. G., De Crescenzo, F., Cipriani, A., Schmidt, U., Treasure, J., & Zipfel, S. (2021, Mar 2021). Comparative efficacy and acceptability of psychological interventions for the treatment of adult outpatients with anorexia nervosa: a systematic review and network meta-analysis. *The Lancet Psychiatry, 8*(3), 215-224.

(40) Turton, R., Bruidegom, K., Cardi, V., Hirsch, C. R., & Treasure, J. (2016). Novel methods to help develop healthier eating habits for eating and weight disorders: A systematic review and meta-analysis [Review]. *Neuroscience and Biobehavioral Reviews, 61*, 132-155.

(41) van den Berg, E., Houtzager, L., de Vos, J., Daemen, I., Katsaragaki, G., Karyotaki, E., Cuijpers, P., & Dekker, J. (2019). Meta-analysis on the efficacy of psychological treatments for anorexia nervosa [Review]. *European Eating Disorders Review, 27*(4), 331-351.

(42) Zeeck, A., Herpertz-Dahlmann, B., Friederich, H. C., Brockmeyer, T., Resmark, G., Hagenah, U., Ehrlich, S., Cuntz, U., Zipfel, S., & Hartmann, A. (2018). Psychotherapeutic treatment for anorexia nervosa: A systematic review and network meta-analysis. *Frontiers in Psychiatry, 9*, Article 158.

Less than three cognitive remediation studies included

(43) Couturier, J., Pellegrini, D., Miller, C., Bhatnagar, N., Boachie, A., Bourret, K., Brouwers, M., Coelho, J. S., Dimitropoulos, G., Findlay, S., Ford, C., Geller, J., Grewal, S., Gusella, J., Isserlin, L., Jericho, M., Johnson, N., Katzman, D. K., Kimber, M., Lafrance, A., Leclerc, A., Loewen, R., Loewen, T., McVey, G., Norris, M., Pilon, D., Preskow, W., Spettigue, W., Steinegger, C., Waite, E., & Webb, C. (2021, Apr). The COVID-19 pandemic and eating disorders in

children, adolescents, and emerging adults: virtual care recommendations from the Canadian consensus panel during COVID-19 and beyond. *Journal of Eating Disorders, 9*(1), Article 46.

(44) Hay, P. J., Claudino, A. M., Touyz, S., & Abd Elbaky, G. (2015, Jul 27). Individual psychological therapy in the outpatient treatment of adults with anorexia nervosa. *Cochrane Database of Systematic Reviews, 2015*(7), Cd003909.

(45) Hay, P. J., Sud, R., & Touyz, S. (2012, Dec 2012). Treatment for severe and enduring anorexia nervosa: A review. *Australian and New Zealand Journal of Psychiatry, 46*(12), 1136-1144.

Less than three studies specifying anorexia nervose as population

(46) Spattini, L., Rioli, G., Longo, F., Ferrari, S., & Galeazzi, G. M. (2017). An update on Current Clinical Management of Eating Disorders. *Minerva Psichiatrica, 58*(1), 54-69.

Review explored cognitive processing and not interventions

(47) Keegan, E., Tchanturia, K., & Wade, T. D. (2021, Mar). Central coherence and set-shifting between nonunderweight eating disorders and anorexia nervosa: A systematic review and meta-analysis. *International Journal of Eating Disorders, 54*(3), 229-243.

(48) Martinez, G., Cook-Darzens, S., Chaste, P., Mouren, M. C., & Doyen, C. (2014, Apr). Anorexia nervosa in the light of neurocognitive functioning: New theoretical and therapeutic perspectives. *Encephale-Revue De Psychiatrie Clinique Biologique Et Therapeutique, 40*(2), 160-167.

(49) Miles, S., Gnatt, I., Phillipou, A., & Nedeljkovic, M. (2020, Nov). Cognitive flexibility in acute anorexia nervosa and after recovery: A systematic review. *Clinical Psychology Review, 81*, Article 101905.

Appendix C. SWAT participant instructions.

Word Task

Instructions: This task involves reading a number of statements that are open to various meanings and/or interpretations. This means that the statements are ambiguous.

You are asked to provide a **single word** that you think **resolves the ambiguity of the statement**. This word should be closely related to the statement.

Try not to spend too long on one question. Your answer should be the **first word that comes to mind**. Once you have inputted the word, you will then be able to move to the next question.

The word should **create one possible meaning** and/or interpretation of the statement.

There are **no right or wrong answers**.

There are **120 statements altogether**. Please see the examples below:

Example statement 1:

Your financial situation has changed.

Possible one-word responses include: '**better**' or '**worse**'

Example statement 2:

You plan to surprise your partner by buying them a watch. You look on the internet to see whether there are any watches within your budget.

Possible one-word responses include: '**affordable**' or '**pricey**'

Example statement 3:

Your performance in the play was commented on by the other actors. You think about what they said.

Possible one-word responses include: '**hurtful**' or '**encouraging**'

Please note, you will be prompted to take a break halfway through completing the task.

Click the arrow to see the first statement and start the task...

You are halfway through the 'Word Task'

Please take a short break…

The instructions for completing the task have been repeated below.

Instructions: This task involves reading a number of statements that are open to various meanings and/or interpretations. This means that the statements are ambiguous.

You are asked to provide a **single word** that you think **resolves the ambiguity of the statement**. This word should be closely related to the statement.

Try not to spend too long on one question. Your answer should be the **first word that comes to mind**. Once you have inputted the word, you will then be able to move to the next question.

The word should **create one possible meaning** and/or interpretation of the statement.

There are **no right or wrong answers**.

There are **120 statements altogether**. Please see the examples below:

Example statement 1:

Your financial situation has changed.

Possible one-word responses include: '**better**' or '**worse**'

Example statement 2:

You plan to surprise your partner by buying them a watch. You look on the internet to see whether there are any watches within your budget.

Possible one-word responses include: '**affordable**' or '**pricey**'

Example statement 3:

Your performance in the play was commented on by the other actors. You think about what they said.

Possible one-word responses include: '**hurtful**' or '**encouraging**'

Click the arrow to proceed with the task…

Appendix D. Ambiguous statements list – 120 items.

Item number	Worry domain	Statement
1	lack of confidence	Some important people are visiting your office and you are asked to present a project to them. You are aware that a lot depends on your presentation and how it goes.
2	relationship	You notice your partner's tone as they comment on how you have handled a difficult situation recently.
3	finance	You get a property tax bill by post. As you see the amount of the total bill, you think about the money in your account and the chance of paying it all at once.
4	relationship	You are completing a group project, but one of the team members is not pulling their weight. You decide to talk about this, and as you start to speak, you know what this conversation will be like.
5	relationship	You are sitting in the café when your date comes in. You notice their expression when they see you, and you know what their first impression of you is.
6	lack of confidence	You have a problem with your gas bill so you phone customer services about a refund. You explain the situation and you know what the outcome will be.
7	performance	There will be a layoff of some roles in your office. When you meet with HR, they tell you about your position.
8	performance	You organise a work party which goes well apart from running low on snacks. Your boss tells you what they think about your organisation skills.
9	harm or illness	The value on the blood pressure meter indicates your blood pressure.
10	harm or illness	You go to the doctor for a check-up and she tells you about your health.
11	performance	You have been working hard on a project at work. You think about the amount of work you still need to do and how it will be to meet the deadline.
12	finance	Your daughter's school is organising a trip. She is desperate to go and you wonder how much the school will charge.
13	relationship	You arranged to have dinner with a new date a week ago. You have not been in contact much since. When you arrive at the restaurant they are not there and you know the reason.
14	finance	You applied for a loan and you receive a call from a loan officer.
15	harm or illness	You cook your friends a meal. After they have left, you notice that the milk you used was a week out of date. You ring them to check how they feel.
16	general	Your boiler has broken and you are told what it will take to find and fix the problems.
17	harm or illness	You hear footsteps running towards you. As they get closer, you turn around and see someone.
18	harm or illness	You are playing throw and catch with your dog. You notice that he is breathing very quickly and you think it is a sign.

19	relationship	At your first meeting in your new job, you compare what everyone is wearing to your own outfit and realise what your new colleagues will think about you.
20	relationship	Your partner is away on a trip and has not answered your call since last night.
21	relationship	You are not quite sure if the birthday present you picked was something your partner would like. As your partner unwraps the gift, you know what your choice was like.
22	finance	You just moved into an unfurnished apartment. When you search on the internet to see what furniture is within your budget, you're surprised.
23	performance	You ask your teacher a question, and by her tone, you know what she thinks of your question.
24	general	You made a plan to exercise regularly. After a month, you think about how your plan is going.
25	harm or illness	Your father's health is surprising given his age.
26	performance	You are on a training course working in a group and the course leader asks you a question. As you give your answer, you realise what the rest of the group thinks about you.
27	finance	Your child needs a new laptop. Looking at your bank balance, you know what the likelihood is that you are able to contribute some money towards it.
28	finance	You look at your bank statement and you are surprised at the total.
29	performance	You register for a course. Looking at the reading list for the first module, you think about the time you have to get through it.
30	aimless future	You are about to start a new job. You think about the challenges that you may face in your new role.
31	performance	Your company gives a pay bonus to workers who exceed their monthly targets. You search for your name on the list.
32	performance	You take part in a half marathon. After the race is over, you are surprised at your time.
33	relationship	You are having dinner with your whole family, and you know what it will be like.
34	relationship	You take some items back to the shop. The assistant agrees to give you a refund and you notice her attitude.
35	performance	You start a new job and realise something about the time allocated for you to learn to use their new computer software before you have to start working.
36	performance	You complete a test and as you look around the room, you wonder what it means if you are the last to finish.
37	aimless future	There will be a role change in your office, and you have a meeting with HR.
38	performance	This is the first time you are leading a meeting in your office. When you start talking, you are aware of how you sound.
39	performance	You make a few errors in the driving test. At the end, the expression on the assessor's face indicates the result.

40	general	You are waiting at a bus stop to go to your Doctor's appointment, and when you see the time the next bus is due, you realise when you will get to your appointment.
41	performance	Your manager says that the quality of your work has changed.
42	performance	You've just started a new job and try to be useful. Today your boss wants to talk about your work productivity, and you know what she thinks.
43	performance	You enter a photography contest and you think about the quality of your photographs.
44	harm or illness	You are at the beach with your family. One of your children is in the water and is trying to get your attention.
45	relationship	Your friend is very keen on skating and invites you to go with her. You are not very good at skating so you plan to turn down her invitation, and you know how she will feel.
46	aimless future	You read a report on people's level of contentment and life satisfaction, and you think about your own life and the future.
47	general	You can see the bus at the bus stop and you are a little way off. You start to run and wave at the driver, and you know what the bus will do.
48	performance	The class is silent throughout your presentation and you notice how the audience looks at you.
49	lack of confidence	You have been working hard in your spare time to build a garden patio. You have never done this before and as you finish and look at the patio, you think about your efforts.
50	relationship	Your friend opens your gift. You can tell how he feels by his face.
51	performance	You hear that your supervisor is looking for you, you think about your work recently and know what they will say.
52	finance	You make a quick decision to buy a laptop. When you see the price of the same laptop in another store, you think about your decision.
53	harm or illness	The doctor examines your X ray film, checks a dark spot and tells you what it is.
54	relationship	You make a joke about your friend, and you can tell what he thinks from his expression.
55	aimless future	You feel unsure about your career choice and consider switching career paths, and you think of starting something new.
56	harm or illness	You step on to the icy path and feel yourself slip. You reach out to try to steady yourself.
57	general	You try to quit high calorie food. After a week, you know how your plan is going.
58	aimless future	As you think about being a year older, you think about the year ahead and consider how you are feeling.
59	performance	You applied for a scholarship from a university and are waiting for the result. Today you receive the reply.
60	aimless future	You think about the number of life goals you are likely to achieve next year.

61	finance	You have been saving for a new phone for a while. Today you check how much you've saved and think about buying a new phone now.
62	relationship	Today is your first day in your new job. As you walk into the office, your new colleagues talk to you.
63	performance	Some other writers comment on the magazine article you wrote.
64	performance	Your examiner looks at the floor during your presentation.
65	general	You had to leave the house in a rush this morning. When you return home in the evening, you check your kitchen faucet.
66	relationship	You've just transferred to a new department. You want to join your new classmates at lunch time. As you walk up to them, you see their expressions.
67	relationship	You overhear your manager saying something about you to another colleague.
68	relationship	Your friend asks your opinion on a relationship issue. You give him some advice and you can tell what he thinks of it.
69	performance	You started an internship a few weeks ago. You wonder what your supervisor thinks about your ability.
70	relationship	You are meeting your partner's friends for the first time tonight. As you chat with them, you can tell what they think of your conversation.
71	finance	Your computer has broken, so you go to the nearest electrical store to check out the cost of replacing it. You think about your bank balance as you see the prices, and know whether you can afford anything new.
72	performance	Your new supervisor calls a meeting and asks each of you in turn to say a little about yourself and your area of expertise. After you have finished speaking, you are aware of how you came across.
73	general	You arrive ten minutes late for your Doctor's appointment. You ask the receptionist if you can still see the doctor and she quickly replies.
74	relationship	Your performance in the play was commented on by the other actors.
75	finance	Seeing the price of washing machines in the store makes you realise whether you can buy a new one.
76	performance	You are having an interview for some funding that you applied for. You can tell what the committee thinks of your experience when you present your resume.
77	lack of confidence	You picked an outfit for today's party very carefully. When you are at the party, you can tell from others' expressions how you look.
78	general	The fitness instructor at the gym tests your muscle strength and provides feedback about your physical condition.
79	lack of confidence	You take your friend to a restaurant and recommend some of your favourite dishes to him. You know what he thinks of the taste by the expression on his face when he starts eating.
80	lack of confidence	You left the house in a hurry this morning. At the end of the day, you are on your way home, and wonder whether you left the door open.

81	harm or Illness	You are hiking up a mountain. At the halfway point, you notice that your face is very sweaty, and you think about why this might be.
82	relationship	You buy a wedding gift for your friend. You can tell what she thinks of it from her facial expression.
83	performance	You spent a lot of time on a tough course this term. As you login to see your final grade, you think about the possible outcome of all your efforts.
84	performance	You have been in a new work environment for two months, and you reflect on how you are doing in the role.
85	aimless future	It's New Year's eve. As you think about what the year ahead might bring, you notice how you are feeling.
86	aimless future	It is your last day in your current job. You consider your future and think about what your next career move will be.
87	lack of confidence	You have a very formal job interview today and spent hours picking out an outfit. As you enter the room, you can tell by the way the interview panel look at you what they think about your outfit.
88	harm or illness	You baby sit for your neighbour and put their daughter to bed. An hour later you go in to check on her and notice something.
89	performance	You are about to sit an assessment following a trial period in a new job. As you sit down to take the test you are quite sure of how it will go.
90	performance	You think about how your new job has changed your life.
91	harm or illness	You have a blood test as part of a routine health check. When the results come back you see how you are doing.
92	harm or illness	You hear the captain make an announcement on the plane. You notice the way his voice sounds.
93	aimless future	As you reflect on your career goals and how much progress you have made so far, you realise how you are doing.
94	aimless future	Thinking about your future in the company, you realise your chances of promotion.
95	finance	Your company is making a few changes to their salary structure. You think about your future earnings compared to now.
96	aimless future	You wrote a list about what you hope to achieve this year. As you review what you have and haven't done, you reflect on how you feel.
97	finance	You get your electricity bill and when you check the total amount, you realise whether you can pay it all at once.
98	performance	You overhear your classmates discussing your contribution to the presentation.
99	finance	You have received a letter from the bank notifying you of changes to your interest rate.
100	performance	You are preparing for your annual work appraisal. You think about what is required to pass.
101	performance	You participate in an art competition. When you see the other entrant's work, you think about your own and know the chances of you winning the prize.

102	relationship	You are having a reunion dinner with your old classmates. You think about what it will be like to see them again.
103	performance	Your boss wants to meet with you, and you think about what they will say.
104	relationship	You are unsure whether you might have offended your friend. When you next speak to her, you can tell how she feels by the tone of her voice.
105	performance	You are preparing for a test as part of a work diploma. You think about how you are likely to do.
106	finance	You are looking for a room in a apartment share. You find one in a suitable area and find the rent is contrary to your expectations.
107	relationship	Your relationship with your parents has changed.
108	aimless future	You are graduating next week and you think about the future ahead.
109	general	You can't find your purse, so you call your family to see if you left it at home. After a while your family calls back.
110	performance	You have a difficult task today. At the end of the day, a senior colleague reviews all the parts that you have done.
111	relationship	Your have promised your partner to pick them up from work today. The traffic is bad and you are going to be late. You know how your partner will react to you being late.
112	finance	You are looking for a birthday present for your mother. You see one that your mother will like, and you check the price tag.
113	performance	You worked really hard on your final project. As you receive the comments, you realise something about all your efforts.
114	relationship	You have been promoted to head the team. You see your new team members are chatting and want to join them. As you walk up to them, you can tell by the look on their faces whether they want you to approach.
115	relationship	You and your friends are trying to choose a film for a movie night. As you suggest a film which you really enjoyed, you see them exchange a look and you know what they think of your choice.
116	finance	Looking at your bank balance at the end of the month, you can see how you have managed your finances.
117	finance	You have been trying to save for your daughter's university fees. Before her graduation from high school, you check your bank statement and realise whether you can contribute towards it.
118	performance	Your colleague asks your opinion about a difficult issue. You tell them your thoughts, and by the expression on their face, you realise what they think of your advice.
119	performance	You recently applied for a work secondment to an interesting project that others are also applying to. You think of your recent performance at work and realise your chance of getting it.
120	performance	There will be major changes at your company over the next 3 years, and you think about your possible future role.

Appendix E. Participant information sheet.

Participant Information Sheet

The impact of mood and thinking style on performance on a sentence and word-based task

Researchers: Libby Tindle, Dr Charlotte Krahé, Dr Frances Meeten

You have been invited to participate in our research study at the University of Liverpool. Before you decide to participate, it is important for you to understand why the research is being done and what participation involves. We have made a list of relevant questions and answered them below. Please take time to read the information carefully. Our contact details have been provided at the end of this page, should you have any concerns or further questions. We would like to stress that you do not have to accept this invitation and should only agree to take part if you want to.

What is the research project title?
The impact of mood and thinking style on performance on a sentence and word-based task.

Why is this research being done?
Worry is a common experience, and we all worry from time to time. It can be understood as a series of thoughts and images about the future, involving uncertain and possibly negative outcomes. Worry is a style of thinking and can be linked to anxious mood. People vary in how much they worry, and we are interested in the impact of mood and thinking style on performance on a sentence and word-based task.

Do I have to take part?
In this study, we are looking to recruit people from the general population. There is no expectation that you should take part. If you do decide to take part, you should indicate your agreement in the consent form on the next page. You can withdraw from this study at any time and you do not have to provide a reason for this.

What will happen if I take part?
There are a series of questionnaires to be completed at 2 separate time points, which are to be completed 2 weeks apart. They will take approximately 60 minutes each. You will first be asked to complete a short questionnaire about who is participating (e.g., your age), followed by 4 questionnaires relating to your mood (anxiety and depression). You will then be presented with a series of statements that are open to various meanings and/or interpretations. This means that the statements are ambiguous. You will be asked to provide a single word, which you think resolves the ambiguity of the statement (and is the most closely related word). Instructions and examples of generating the one-word response will be provided, in order to demonstrate what is required on the task.

We recognise that questions related to worry, depression and anxiety can be sensitive. We have therefore provided the web address for the Anxiety and Depression Association of America (ADAA) at the end of the questionnaires (see https://adaa.org).

ADAA is an international non-profit organisation that focusses on the prevention and treatment of mental health difficulties. ADAA helps people find treatment, resources, and support.

Who is eligible to take part?
We are recruiting adults aged 18-65 years who are resident in the US. English fluency and English as a native language is a requirement due to the linguistic nature of the tasks included in the study (i.e., involving the use of abstract sentences and language). This is an online study and requires participants to be familiar with the use of computers (i.e., using a keyboard and mouse to navigate).

What will I have to do if I am eligible and want to take part
If you decide that you would like to participate in the study, you will be provided with a link to the study using Qualtrics (where you are accessing this information) via Amazon Mechanical Turk (MTurk). Once you have read the participant information, you will then be required to give informed consent. Following completion of the questionnaires (at each time point), you will be provided with a unique 10-digit reference number. You will be asked to paste the number into the MTurk window where you accessed this link, to allow for compensation of your time.

What will happen after I have completed the questionnaires at both time points?
At the end of the questionnaires at the second time point, you will be provided with participant debrief information. The contact details of the researchers will also be provided, should you want to ask any questions or raise concerns.

Are there any disadvantages to taking part?
The study asks questions related to worry, depression and anxiety, which can be sensitive areas. If you feel uncomfortable and do not want to continue, you can end the task at any point and withdraw from the study.

Are there any advantages to taking part?
Whilst there are no immediate benefits for those participating in the study, it is hoped that it will contribute to the evidence base concerned with improving the effectiveness and efficiency of interventions relating to excessive worry and anxiety.

Will I be compensated for my participation?
Participants will be compensated with $4.80 (US Dollars) following completion of the questionnaires, at each time point of the study.

Will my taking part in this project be kept confidential?
All information will be kept strictly confidential, and data will be stored securely and anonymously, in accordance with the University of Liverpool's Research Data Management policy. All data will be provided in anonymised form, and only the researchers directly involved in the study will have access to this.

Will I be recorded, and how will the recorded media be used?
You will not be recorded in any way other than your anonymous input into the questionnaires and demographic information.

How will my data be used?
The University processes personal data as part of its research and teaching activities in accordance with the lawful basis of 'public task', and in accordance with the University's purpose of advancing education, learning and research for the public benefit.

Under UK data protection legislation, the University acts as the Data Controller for personal data collected as part of the University's research. The Principal Investigator acts as the Data Processor for this study, and any queries relating to the handling of your personal data can be sent to Dr Charlotte Krahé (see contact details below).

Further information on how your data will be used can be found below:

How will my data be collected?
Your data will be collected via a program called Qualtrics.
How will my data be stored?
Your data will be stored in accordance with the University of Liverpool's Research Data Management policy.
How long will my data be stored for?
Up to 10 years.
What measures are in place to protect the security and confidentiality of my data?
All electronic data and information will be protected with security passwords. All data generated from the main study will be stored in the University of Liverpool's secure data storage facilities (i.e., Active DataStore).
Will my data be anonymised?
All data will be 'pseudo-anonymised' up until the point at which it is analysed, which will be 2 weeks after closing the second link. Should you want to withdraw any data provided, you can contact the researchers and provide your MTurk ID number, and we will be able to match you to your data. We will not ask you to provide any other identifiable information or contact details (e.g., name, email address). Your data will be fully anonymised at the point of data analysis, which means that results may only be withdrawn prior to this.
How will my data be used?
The data will only be used for the current study.
Who will have access to my data?
Only those directly involved in the research will have access to your data.
How will my data be destroyed?
Data will be destroyed in accordance with the University of Liverpool's Research Data Management policy, which will remain the responsibility of Dr Charlotte Krahé (Data Custodian).

Who has approved the project?
This study was approved by the DClinPsychol Research Committee, which includes reviewers by two readers, a member of the Liverpool Expert by Experience team, as well as an NHS representative and the Clinical Research Governance team. It has received ethical approval from the University of Liverpool Central Research Ethics Committee.

What will happen to the results of the study?
We will write a summary of our findings and it is hoped that the research will be published within a journal interested in understanding emotional and behavioural disorders, their prevention and treatment. You will not be identified in any report or publication.

What will happen if I want to stop taking part?
You can end your participation during the study at any time, without explanation.

What if there is a problem?
If there is a problem, please let us know by contacting Dr Charlotte Krahé (see contact details below) and we will try to help. If you remain unhappy or have a complaint that you feel you cannot come to us with then you should contact the Research Ethics and Integrity Office at ethics@liv.ac.uk. When contacting the Research Ethics and Integrity Office, please provide details of the name or description of the study (so that it can be identified), the researchers involved, and the details of the complaint you wish to make.

The University strives to maintain the highest standards of rigour in the processing of your data. However, if you have any concerns about the way in which the University processes your personal data, it is important that you are aware of your right to lodge a complaint with the Information Commissioner's Office by calling 0303 123 1113.

Who can I contact if I have further questions?
If you have any further questions, please contact us. Our details can be found below:

Principal Investigator
Dr Charlotte Krahé
Lecturer in Clinical Psychology
Department of Psychological Sciences
University of Liverpool
Whelan Building
Liverpool
L69 3GB
Tel: 0151 794 5530
Email: Charlotte.Krahé@liverpool.ac.uk

Doctoral Project Researcher
Libby Tindle
Trainee Clinical Psychologist
Department of Psychological Sciences
University of Liverpool
Whelan Building
Liverpool
L69 3GB
Tel: 0151 794 5530
Email: Libby.Tindle@liverpool.ac.uk

Thank you for reading this

Appendix F. Participant consent form.

Consent to Participate

Study title: The impact of mood and thinking style on performance on a sentence and word-based task

Researchers: Libby Tindle, Dr Charlotte Krahé, Dr Frances Meeten

1. I confirm that I have read and have understood the information sheet dated 07/05/2021 for the above study.

- Please select

2. I understand that taking part in the study involves completing a series of questionnaires at 2 separate time points, which are to be 2-weeks apart.

- Please select

3. I understand that I must only take part if:
  - I am aged between 18-65 years
  - I am fluent in English
  - English is my native language
  - I am a resident in the US
  - I am familiar with the use of computers

- Please select

4. I understand that the study involves questions relating to worry, depression, and anxiety, which can be sensitive areas.

- Please select

5. I understand that my participation is voluntary and that I am free to stop taking part at any time, without giving any reason, and without my rights being affected.

- Please select

6. I understand that I may withdraw any data I provide, up until the point at which it is analysed, which will be 2 weeks after closing the second link. Your MTurk ID number will allow us to match you to your data. Your data will be fully anonymised at the point of data analysis, which means that results may only be withdrawn prior to this.

- Please select

7. I understand that confidentiality and anonymity will be maintained, and it will not be possible to identify me in any publications. I give permission for members of the research team to have access to my fully anonymised responses.

- Please select

8. I understand that the information I provide will be held securely and in line with data protection requirements at the University of Liverpool.

- Please select

9. I agree to take part in the above study.

- Please select

Appendix G. Demographic questions.

Demographics Questions

Please enter your MTurk worker ID:

Age (in years):

"*drop down box ages 18-65*"

Gender:

Is English your native language?

- Yes
- No

Occupation:

What is your highest level of education?

- High school graduate
- Bachelor
- Master
- Doctoral/PhD
- Other:

Country of residency:

Appendix H. Participant debrief sheet.

Participant Debrief Sheet

The impact of mood and thinking style on performance on a sentence and word-based task

Researchers: Libby Tindle, Dr Charlotte Krahé, Dr Frances Meeten

Thank you for taking the time to participate in our study. We are very thankful for your interest and appreciate your involvement.

What is the full title of the project?
Developing a measure of interpretation bias for worry: a study of reliability and validity.

Is there any further information about the study?
Life is filled with ambiguity, and people seek to make sense of what might be happening to them or how others are behaving. A resolution of ambiguity is often seen to be crucial as a person anticipates a future event, or creates mental representations of the past. This might help them to problem-solve, as well as plan what they are going to do. An example might include giving a speech and then hearing laughter, which could suggest warmth or friendliness towards someone, but also hostility and disagreement.

People vary in how they interpret ambiguous events. Those who experience depression and/or anxiety are more likely than others to interpret ambiguity in a negative way, attend more to emotionally negative cues, and selectively recall negative information. People who experience anxiety often report worrying too much, and/or not being able to stop worrying. Interpretation bias refers to the tendency to draw inferences about ambiguous information in consistent ways.

A number of different measures have been developed to assess interpretation bias, as well as adapted to train a more positive interpretation bias. Yet, few studies have looked at the reliability and validity of the measures used. This project therefore proposes to build on existing interpretation bias methodologies, specifically in relation to worry. We aim to develop material for an interpretation bias task to be used for assessment and training purposes. The objective of the study is to validate a measure of interpretation bias by looking at the resolutions that people have made to a series of ambiguous statements. It is hypothesised that the more positive resolutions generated, the lower the reported anxiety.

How was this tested?
You completed a number of measures relating to worry, anxiety, interpretation bias, and depression. The statement task that was developed assessed how you interpret ambiguous events relating to worry.

Why did I not know all the information before taking part?
In this research we were looking to assess interpretation bias. We wanted participants to respond as naturally as possible, to not affect the outcome of the study.

How will the results be evaluated?
The results will form part of a doctoral thesis at the University of Liverpool.

Why is this study important?
It is hoped that the materials developed can be used in future studies, including as a measure of interpretation bias whereby people are asked to make a relatedness judgement (i.e., supraliminally). Future research may also seek to understand the stages of processing at which people make interpretations (e.g., subliminal training paradigms). This could then have implications for further research into potential novel treatments for anxiety.

What if I no longer want my information to be included in the research?
I may withdraw any data I provide, up until the point at which it is analysed, which will be 2 weeks after closing the second link. Your MTurk ID number will allow us to match you to your data. Your data will be fully anonymised at the point of data analysis, which means that results may only be withdrawn prior to this.

Where can I find relevant support?
Worry is a common experience, and we all worry from time to time; however, we understand that some participants might experience anxiety or depression. Should you feel that you might require support in relation to this, you might want to make an appointment with your Doctor and discuss this further. You might also find the below contact helpful:

Anxiety and Depression Association of America (ADAA)

ADAA is an international non-profit organisation that focusses on the prevention and treatment of mental health difficulties.

ADAA helps people find treatment, resources, and support. Please refer to the below web address for further information:

Website: https://adaa.org

What if I want to know more?
If you are interested in learning more about the study, please contact the Principal Investigator, Dr Charlotte Krahé. See contact details below:

Principal Investigator
Dr Charlotte Krahé
Lecturer in Clinical Psychology
Department of Psychological Sciences
University of Liverpool
Whelan Building
Liverpool
L69 3GB
Tel: 0151 794 5530
Email: Charlotte.Krahé@liverpool.ac.uk

Doctoral Project Researcher
Libby Tindle
Trainee Clinical Psychologist
Department of Psychological Sciences
University of Liverpool
Whelan Building
Liverpool
L69 3GB
Tel: 0151 794 5530
Email: Libby.Tindle@liverpool.ac.uk

Thank you again for your participation

Appendix I. Participant support information.

Support Information

Whilst worry is a common experience and we all worry from time to time, we understand that some participants might experience anxiety or depression. Should you feel that you require support in relation to this, you might want to make an appointment with your Doctor and discuss this further. You might also find the below contact helpful:

Anxiety and Depression Association of America (ADAA)

ADAA is an international non-profit organisation that focusses on the prevention and treatment of mental health difficulties.

ADAA helps people find treatment, resources, and support. Please refer to the below web address for further information:

Website: https://adaa.org

Appendix J. Ethical approval.

UNIVERSITY OF LIVERPOOL

Institute of Population Health Research Ethics Committee

14 June 2021

Dear Dr Krahe,

I am pleased to inform you that the amendment to your study has been approved. Amendment details and conditions of approval can be found below. If applicable, Appendix A contains a list of documents approved by the Committee.

Amendment details

Reference:	6145 (amendment)
Project Title:	An investigation into the underlying cognitive mechanisms of worry
Principal Investigator:	Dr Charlotte Krahe
Co-Investigator(s):	Miss Libby Tindle
Student Investigator(s):	-
Department:	Psychological Sciences
Approval Date:	14/06/2021

The amendment was **APPROVED** subject to the following conditions:

Conditions of approval

Please note: this approval is subject to the University's research restrictions during the pandemic, as laid out on the research ethics webpages. Therefore, wherever possible, research should be conducted via remote means which avoid the need for face-to-face contact with human participants during the pandemic. The process for requesting an exemption to these restrictions is described on the research ethics webpages.

- All serious adverse events must be reported to the Committee (ethics@liv.ac.uk) in accordance with the procedure for reporting adverse events.
- If it is proposed to make further amendments to the study, please create and submit an amendment form within the research ethics system.
- It is the responsibility of the Principal Investigator or Supervisor to inform all the investigators of the terms of the approval.

Kind regards,

Institute of Population Health Research Ethics Committee

iphethics@liverpool.ac.uk

IPH-REC

Appendix K. Frequency table – individual items of the SWAT.

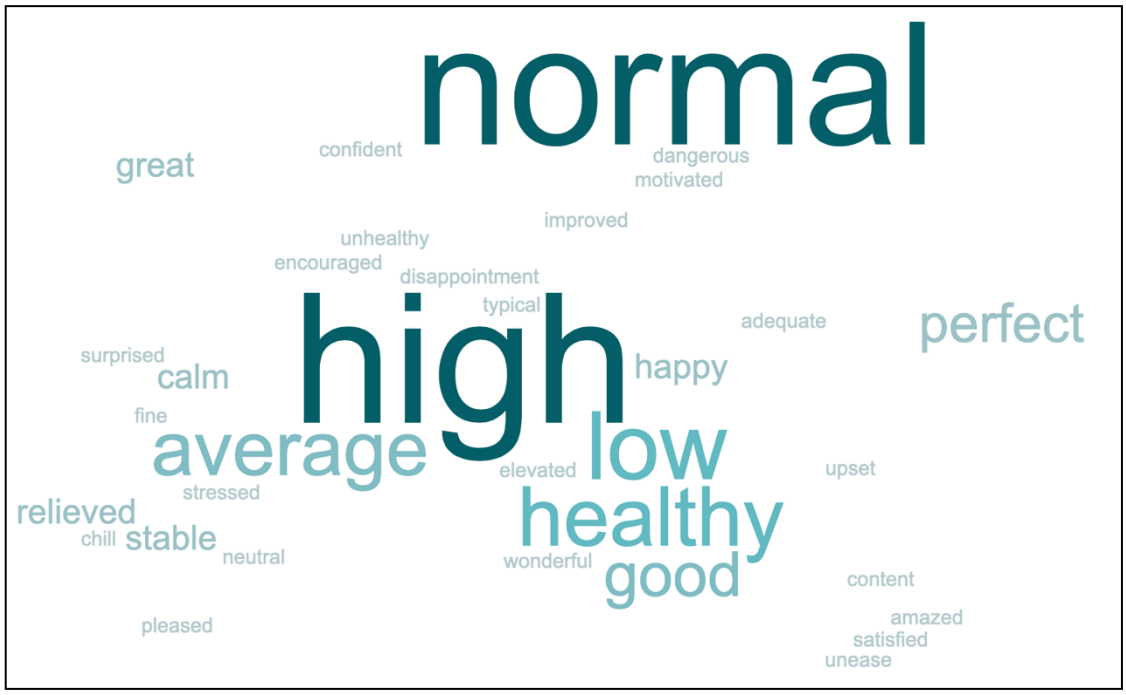
Item number	Positive IB		Negative IB		Un-codable	
	TP1	TP2	TP1	TP2	TP1	TP2
1	43.7%	53.8%	52.1%	43.8%	4.2%	2.3%
2	52.1%	64.6%	44.4%	31.5%	3.5%	3.8%
3	40.1%	37.7%	54.2%	54.6%	5.6%	7.7%
~~4~~	~~18.3%~~	~~18.5%~~	~~78.2%~~	~~80.0%~~	~~3.5%~~	~~1.5%~~
5	71.5%	71.5%	23.9%	26.9%	4.9%	1.5%
6	64.1%	70%	33.8%	26.9	2.1%	1.5%
7	51.4%	62.3%	43.0%	36.9%	5.6%	0.8%
8	69.0%	72.3%	29.6%	27.7%	1.4%	0%
9	55.6%	69.2%	38.0%	29.2%	6.3%	1.5%
10	71.1%	73.1%	23.2%	26.2%	5.6%	0.8%
11	26.8%	36.9%	69.%	60.0%	4.2%	3.1%
12	28.9%	33.8%	54.9%	56.9%	16.2%	9.2%
13	33.8%	39.2%	62.7%	57.7%	3.5%	3.1%
14	64.8%	69.2%	31.7%	30.0%	3.5%	0.8%
15	27.5%	35.4%	70.4%	63.8%	2.1%	0.8%
~~16~~	~~16.2%~~	~~19.2%~~	~~76.8%~~	~~75.4%~~	~~7.0%~~	~~5.4%~~
17	52.1%	49.2%	47.2%	46.9%	0.7%	3.8%
18	66.2%	66.2%	32.4%	33.8%	1.4%	0%
19	62.0%	60.8%	35.9%	36.9%	2.1%	2.3%
20	28.9%	34.6%	69.7%	65.4%	1.4%	0%
21	78.2%	76.2%	19.0%	19.2%	2.8%	4.6%
22	71.1%	76.2%	23.2%	22.3%	5.6%	1.5%
23	45.1%	48.5%	50.7%	50.0%	4.2%	1.5%
24	70.4%	74.6%	27.5%	23.1%	2.1%	2.3%
25	88.0%	86.2%	7.0%	10.8%	4.9%	3.1%
26	62.7%	69.2%	34.5%	30.0%	2.8%	0.8%
27	64.1%	64.6%	34.5%	35.4%	1.4%	0%
28	56.3%	69.2%	32.4%	27.7%	11.3%	3.1%
29	35.9%	41.5%	57.7%	50.8%	6.3%	7.7%
30	43.0%	42.3%	44.4%	50.0%	12.7%	7.7%
~~31~~	~~69%~~	~~76.2%~~	~~24.6%~~	~~16.9%~~	~~6.3%~~	~~6.9%~~
~~32~~	~~86.6%~~	~~78.5%~~	~~9.9%~~	~~14.6%~~	~~3.5%~~	~~6.9%~~
33	57.0%	61.5%	38.0%	36.2%	4.9%	2.3%
34	41.5%	53.8%	54.9%	45.4%	3.5%	0.8%
35	33.8%	41.5%	61.3%	54.6%	4.9%	3.8%
36	38.7%	47.7%	57.0%	50.8%	4.2%	1.5%
37	59.2%	55.4%	32.4%	30.8%	8.5%	13.8%
38	36.6%	41.5%	61.3%	57.7%	2.1%	0.8%
39	36.6%	43.1%	59.2%	53.1%	4.2%	3.8%
40	42.3%	33.8%	54.9%	60.8%	2.8%	5.4%
41	56.3%	71.5%	38.7%	26.9%	4.9%	1.5%
42	74.6%	76.2%	21.8%	22.3%	3.5%	1.5%
43	71.8%	73.8%	25.4%	23.8%	2.8%	2.3%
44	60.6%	52.3%	37.3%	43.8%	2.1%	3.1%
~~45~~	~~7.0%~~	~~10%~~	~~90.1%~~	~~87.7%~~	~~2.8%~~	~~2.3%~~

46	62.0%	69.2%	28.9%	25.4%	9.2%	5.4%
47	57.7%	60.0%	36.6%	35.4%	5.6%	4.6%
48	48.6%	59.2%	46.5%	37.7%	4.9%	3.1%
~~49~~	~~84.5%~~	~~85.4%~~	~~12.7%~~	~~11.5%~~	~~2.8%~~	~~3.1%~~
~~50~~	~~85.9%~~	~~93.1%~~	~~12.7%~~	~~6.9%~~	~~1.4%~~	~~0%~~
51	66.2%	70.8%	26.1%	24.6%	7.7%	4.6%
52	38.0%	43.8%	47.2%	47.7%	14.8%	8.5%
53	30.3%	36.2%	62.7%	60.0%	7.0%	3.8%
54	45.1%	43.8%	52.1%	55.4%	2.8%	0.8%
55	54.2%	57.7%	34.5%	37.7%	11.3%	4.6%
56	29.6%	33.1%	63.4%	64.6%	7.0%	2.3%
57	62.7%	64.6%	35.9%	32.3%	1.4%	3.1%
58	64.8%	59.2%	30.3%	32.3%	4.9%	8.5%
59	65.5%	73.1%	31.7%	22.3%	2.8%	4.6%
60	61.3%	69.2%	28.2%	27.7%	10.6%	3.1%
61	73.9%	72.3%	21.8%	23.8%	4.2%	3.8%
~~62~~	~~87.3%~~	~~85.4%~~	~~9.9%~~	~~11.5%~~	~~2.8%~~	~~3.1%~~
63	66.9%	69.2%	28.9%	28.5%	4.2%	2.3%
64	16.9%	19.2%	75.4%	77.7%	7.7%	3.1%
65	40.1%	50.0%	53.5%	44.6%	6.3%	5.4%
66	73.2%	70.0%	21.1%	26.2%	5.6%	3.8%
67	57.0%	57.7%	35.3%	33.8%	7.7%	8.5%
68	71.1%	72.3%	27.5%	25.4%	1.4%	2.3%
69	69.7%	73.8%	20.4%	22.3%	9.9%	3.8%
~~70~~	~~73.2%~~	~~79.2%~~	~~26.1%~~	~~18.5%~~	~~07%~~	~~2.3%~~
71	45.8%	46.9%	48.6%	48.5%	5.6%	4.6%
72	57.0%	61.5%	43.0%	38.5%	0.0%	0%
73	62.7%	56.9%	37.3%	43.1%	0.0%	0%
~~74~~	~~90.1%~~	~~87.7%~~	~~8.5%~~	~~12.3%~~	~~1.4%~~	~~0%~~
75	45.1%	41.5%	50.0%	53.8%	4.9%	4.6%
~~76~~	~~67.6%~~	~~80.8%~~	~~30.3%~~	~~19.2%~~	~~4.9%~~	~~0%~~
~~77~~	~~76.8%~~	~~79.2%~~	~~20.4%~~	~~20.0%~~	~~2.8%~~	~~0.8%~~
78	62.7%	62.3%	33.1%	37.7%	4.2%	0%
79	73.2%	75.4%	21.8%	23.8%	4.9%	0.8%
80	33.8%	40.8%	60.6%	53.8%	5.6%	5.4%
81	45.1%	47.7%	54.9%	52.3%	0.0%	0%
~~82~~	~~89.4%~~	~~88.5%~~	~~9.2%~~	~~10.8%~~	~~1.4%~~	~~0.8%~~
~~83~~	~~76.1%~~	~~80.0%~~	~~23.2%~~	~~17.7%~~	~~0.7%~~	~~2.3%~~
84	85.9%	83.8%	11.3%	13.1%	2.8%	3.1%
85	63.4%	69.2%	33.1%	29.2%	3.5%	1.5%
~~86~~	~~71.1%~~	~~73.8%~~	~~24.6%~~	~~20.0%~~	~~4.2%~~	~~6.2%~~
87	78.9%	72.3%	19.0%	26.2%	2.1%	1.5%
88	48.6%	64.6%	47.2%	33.1%	4.2%	2.3%
89	77.5%	77.7%	21.5%	20.8%	1.4%	1.5%
~~90~~	~~85.2%~~	~~86.2%~~	~~11.3%~~	~~10.0%~~	~~3.5%~~	~~3.8%~~
~~91~~	~~73.2%~~	~~76.9%~~	~~23.9%~~	~~20.0%~~	~~2.8%~~	~~3.1%~~
92	50.0%	53.1%	47.9%	46.2%	2.1%	0.8%
93	77.5%	78.5%	21.1%	20.8%	1.4%	0.8%
94	69.0%	67.7%	28.9%	30.0%	2.1%	2.3%
95	77.5%	67.7%	21.1%	27.7%	1.4%	4.6%

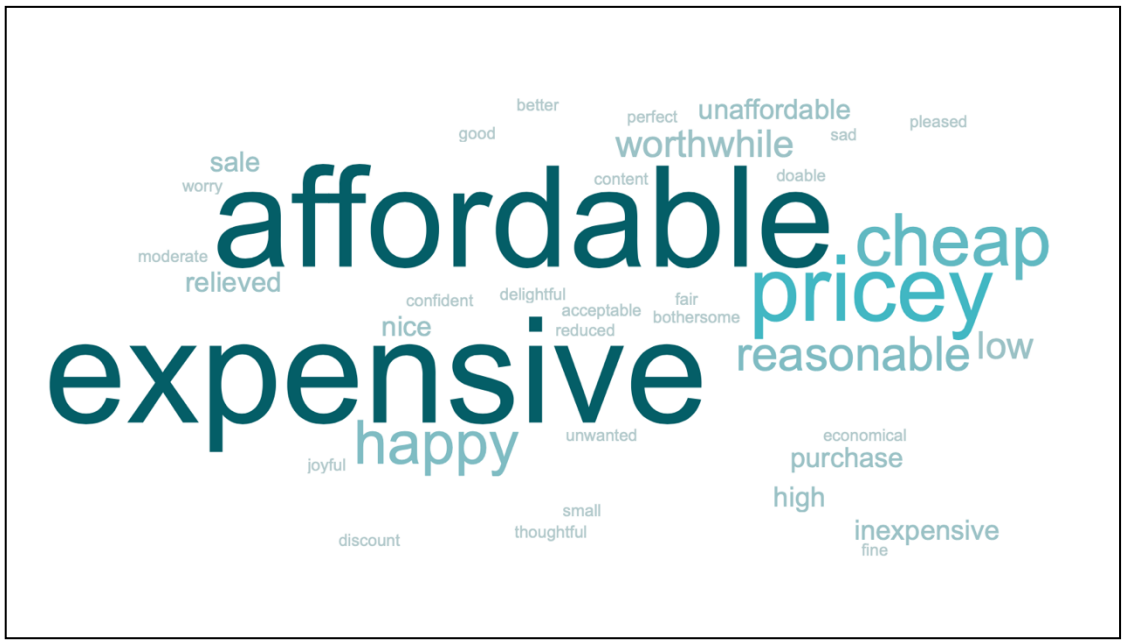
96	58.5%	66.2%	38.0%	30.0%	3.5%	3.8%
97	59.9%	58.5%	35.2%	37.7%	4.9%	5.4%
98	63.4%	69.2%	31.7%	25.4%	4.9%	5.4%
99	43.0%	47.7%	55.6%	49.2%	1.4%	3.1%
100	65.6%	64.6%	30.3%	25.4%	4.2%	10%
101	42.3%	47.7%	54.2%	50.8%	3.5%	1.5%
102	64.1%	60.8%	35.9%	39.2%	0.0%	0%
103	58.5%	61.5%	36.6%	36.9%	4.9%	1.5%
104	35.2%	41.5%	62.0%	57.7%	2.8%	0.8%
~~105~~	~~88.0%~~	~~85.4%~~	~~12.0%~~	~~14.6%~~	~~0.0%~~	~~0%~~
106	52.1%	47.7%	43.7%	49.2%	4.2%	3.1%
107	54.2%	61.5%	40.8%	33.8%	4.9%	4.6%
108	71.1%	70.0%	26.8%	27.7%	2.1%	2.3%
109	73.9%	76.2%	24.6%	21.5%	1.4%	2.3%
~~110~~	~~70.4%~~	~~76.9%~~	~~24.6%~~	~~20.0%~~	~~4.9%~~	~~3.1%~~
111	23.9%	26.2%	73.2%	71.5%	2.8%	2.3%
112	59.2%	60.0%	35.2%	36.2%	5.6%	3.8%
~~113~~	~~85.9%~~	~~85.4%~~	~~13.4%~~	~~13.8%~~	~~0.7%~~	~~0.8%~~
114	67.6%	67.7%	28.9%	29.2%	3.5%	3.1%
115	43.0%	50.8%	47.2%	45.4%	9.9%	3.8%
116	71.1%	76.2%	27.5%	22.3%	1.4%	1.5%
117	61.3%	61.5%	33.1%	37.7%	5.6%	0.8%
118	66.2%	67.7%	31.0%	28.5%	2.8%	3.8%
119	65.5%	72.3%	31.7%	26.9%	2.8%	0.8%
120	62.7%	69.20%	33.1%	24.6%	4.2%	6.2%

Note. IB = interpretation bias; TP1 = time point 2; TP2 = time point 2.

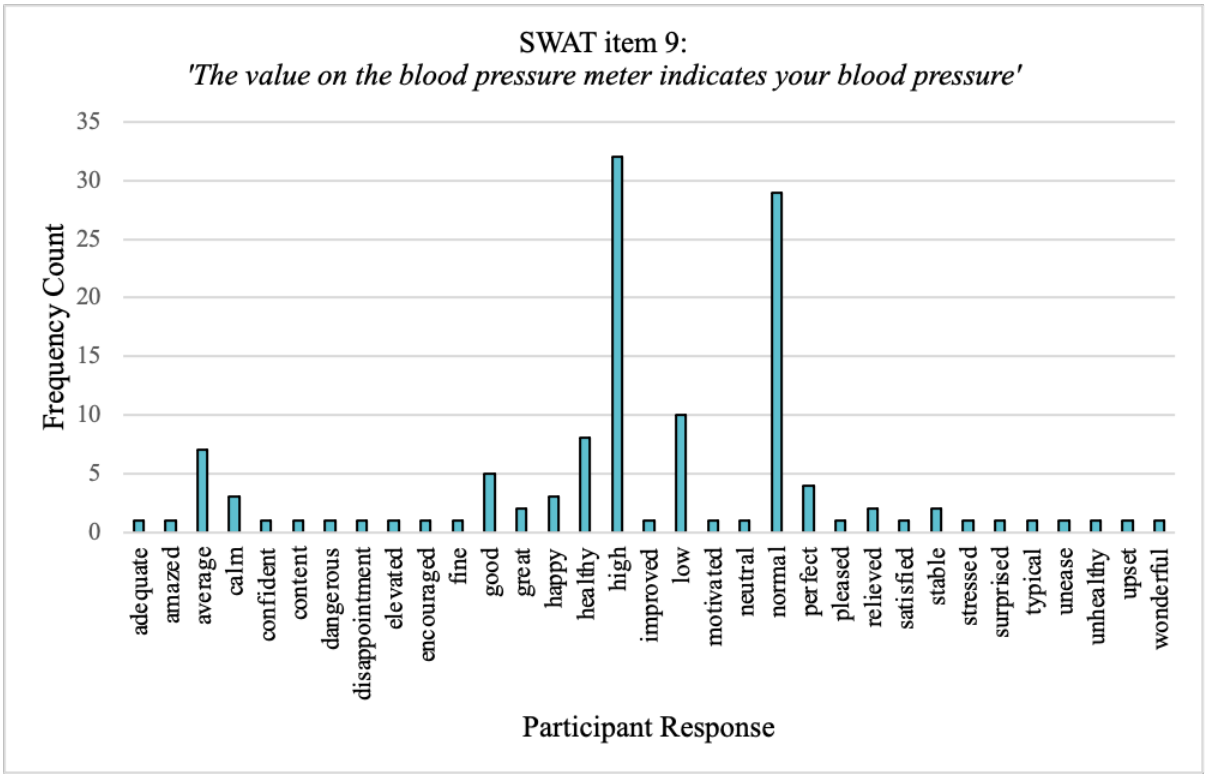
Appendix L. Corresponding results for time point 2 – word clouds and frequency graphs.



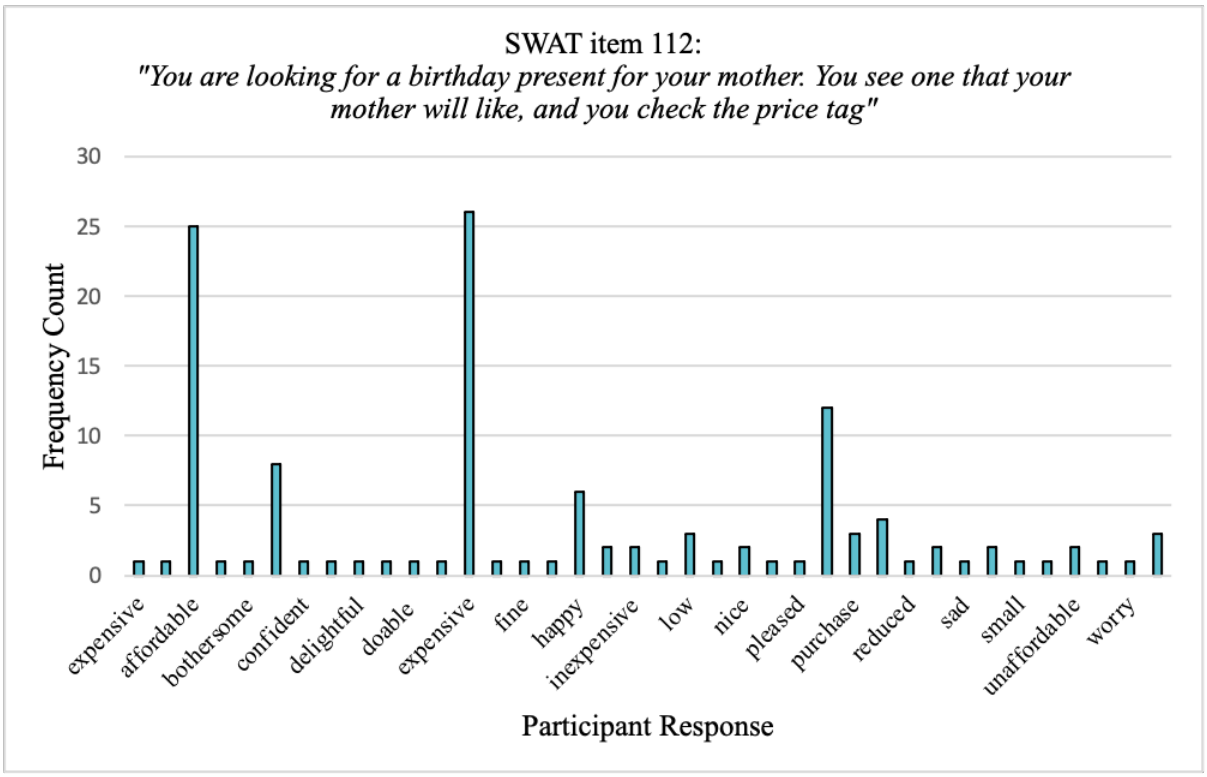
Illustrative word cloud for SWAT item 9 for time point 2.
Note. SWAT item 9: 'The value on the blood pressure meter indicates your blood pressure'.



Illustrative word cloud for SWAT item 112 for time point 2.
Note. SWAT item 112: "You are looking for a birthday present for your mother. You see one that your mother will like, and you check the price tag".



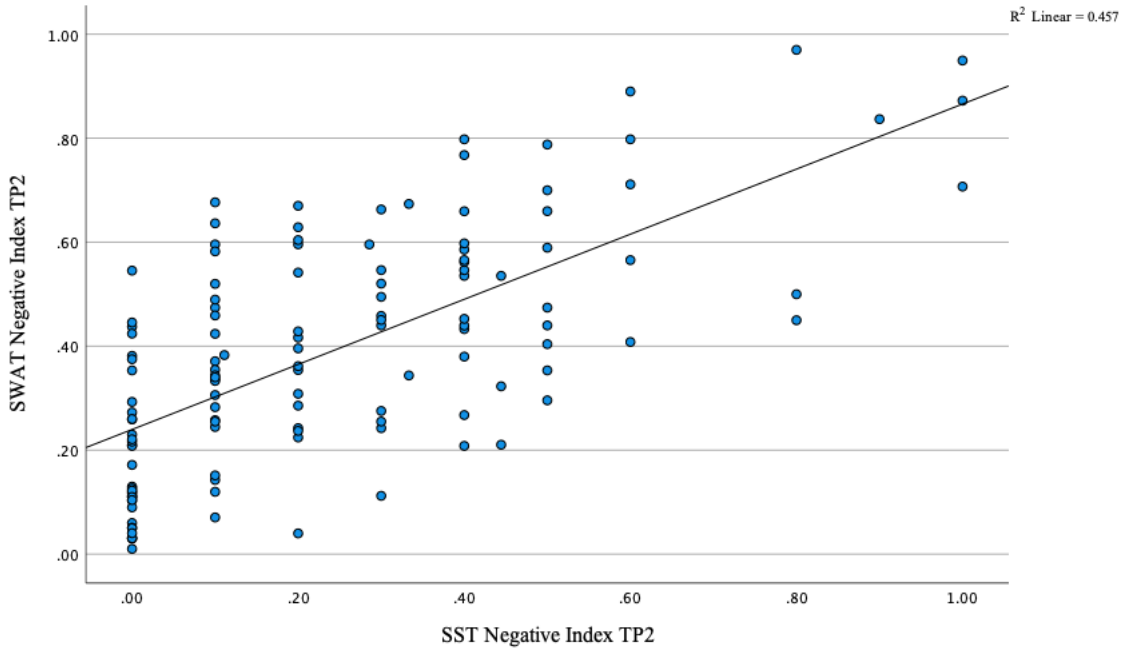
Frequency graph for SWAT item 9 for time point 2.
Note. N = 127. SWAT = Statement Word Association Task. Participant responses presented A-Z.



Frequency graph for SWAT item 112 for time point 2.
Note. N = 124. SWAT = Statement Word Association Task. Participant responses presented A-Z.

Appendix M. Corresponding results for time point 2 – correlations and hierarchical multiple regression.

Association between the SWAT index scores and SST index scores for time point 2.



Note. N=130. SWAT = Statement Word Association Task; SST = Scrambled Sentences Task.

Correlation matrix of the main variables for time point 2.

Spearman's rho Correlations Between Measures					
Measures	**SWAT**	**SST**	**PSWQ**	**GAD-7**	**PHQ-8**
SWAT index					
SST index	.657*				
PSWQ total	.460*	.638*			
GAD-7 total	.368*	.527*	.750*		
PHQ-8 total	.460*	.540*	.599*	.786*	

Note. N=130. SWAT = Statement Word Association Task; SST = Scrambled Sentences Task; PSWQ = Penn State Worry Questionnaire; GAD-7 = Generalised Anxiety Disorder Assessment; PHQ-8 = Patient Health Questionnaire depression scale.
** = correlation is significant at p<.001*

Results from a hierarchical regression assessing predictors of the SWAT for time point 2.

Hierarchical regression with the SWAT as the Criterion Variable

Source	Unstandardised β coefficients	Standard error	Standardised β coefficients	t	Sig.	R^2	Adj. R^2	R^2 change
Step 1						.156	.143	.156
Constant	.312	.024		12.847	<.001			
PHQ-8	.010	.006	.203	1.804	.074			
GAD-7	.011	.005	.227	2.014	.046			
Step 2						.233	.214	.077
Constant	.044	.079		.555	.580			
PHQ-8	.008	.005	.167	1.494	.138			
GAD-7	-.004	.007	-.072	-.527	.599			
PSWQ	.007	.002	.429	3.544	<.001			

Note. N=142. SWAT = Statement Word Association Task; PSWQ = Penn State Worry Questionnaire; GAD-7 = Generalised Anxiety Disorder Assessment; PHQ-8 = Patient Health Questionnaire depression scale.
Step 1 model: $F (2, 127) = 11.738$, $p < .001$, $R^2 = .156$
Step 2 model: $F (1, 126) = 12.561$, $p < .001$, $R^2 = .233$