UNIVERSITY OF
EXETER

LITERATURE REVIEW: Relationships between Facets of Agreeableness and Depression: A Systematic Review

EMPIRICAL PAPER: Using Cognitive Reappraisal and Helping Behaviour to Improve Well-being: A Single-Case Design Study

Submitted by **Tomas Jelinek**, to the University of Exeter

as a thesis for the degree of **Doctor of Clinical Psychology**, May 16th 2019

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature:

COGNITIVE REAPPRAISAL AND WELL-BEING

Contents

COGNITIVE REAPPRAISAL AND WELL-BEING

COGNITIVE REAPPRAISAL AND WELL-BEING

COGNITIVE REAPPRAISAL AND WELL-BEING

List of Tables

List of Figures

UNIVERSITY OF
EXETER

SCHOOL OF PSYCHOLOGY

DOCTORATE IN CLINICAL PSYCHOLOGY

LITERATURE REVIEW

Relationships between Facets of Agreeableness and Depression:

A Systematic Review

Trainee Name: **Tomas Jelinek**

Primary Research Supervisor: **Dr Pia Pechtel**

                                                        Wellcome Trust Research Fellow, University of Exeter

Secondary Research Supervisor: **Dr Nick Moberly**

                                                        Senior Lecturer, University of Exeter

Target Journal: Psychological Bulletin

Word Count: 6506 words (excluding abstract, table of contents, list
                                                        of figures, references, footnotes, appendices)

**Submitted in partial fulfilment of requirements for the Doctorate Degree in Clinical
Psychology, University of Exeter**

Abstract

Depression is the leading cause of disability worldwide. Understanding whether particular personality profiles have a higher risk of developing this disorder could inform prevention and ultimately treatment. Research using the Five Factor Model of personality has highlighted that, whilst links between several personality factors and depression are evident, others remain obscure and require investigations on the level of lower-order facets. One of these factors is Agreeableness, a dimension of interpersonal behaviour and quality of interactions comprising six facets: Trust, Straightforwardness, Altruism, Compliance, Modesty, and Tender-mindedness. To explore possible links between these facets and depression, this review addressed the question: Which facets of Agreeableness are associated with depressive symptoms?

Cross-sectional, correlational and prospective studies assessing associations between personality and depression in adults were identified from multidisciplinary and subject-specific databases published prior to 19th of February 2019 and screened for inclusion according to pre-specified criteria. The systematic literature search yielded 1169 records with 874 non-duplicated results. Screening of 33 full-text papers resulted in nine eligible studies synthesised in this review.

Results yielded weak evidence indicating that Trust and Modesty in depressed adults may be associated with the disorder such that Trust decreases and Modesty increases with symptom severity. Although the quality of included studies was generally poor and causality of these links cannot be established, considering these associations when screening for depression vulnerability may be of high clinical value. Despite

conceptual evidence suggesting links between depression and other facets could exist,

none were found to be statistically significant. Further research is required to test

reported associations with more robust designs.

Keywords: Five Factor Model, personality, depression, systematic review

Introduction

Depression is the leading cause of disability worldwide (World Health Organisation, 2018). This systematic review investigates the potential relationship between depression and lower-order facets of personality based on the Five Factor Model (FFM; Digman, 1997). The FFM conceptualizes personality as hierarchically ordered from specific facets to "Big Five" broad domains (factors) of personality - Neuroticism, Extraversion, Conscientiousness, Agreeableness, and Openness (Goldberg, 1993). Each of these higher-order constructs is composed of several lower-order constructs and this structure is implicit within the many personality inventories that have been designed for personality assessment. A meta-analysis by Kotov, Gamez, Schmidt and Watson (2010) reviewed the associations between higher-order personality factors and psychopathology. Although some links between mood disorders and personality domains were identified, the authors found that several personality domains showed inconsistent relationships with psychopathology. To address this, the authors highlighted the need to investigate these links on a facet level. One of these factors is Agreeableness; a domain that is best understood as a dimension of interpersonal behaviour and quality of interactions (Costa, McCrae, & Dye, 1991). Although Kotov et al. (2010) found Agreeableness to have no relationship with depression, its facets have been theorised to be associated with the disorder.

COGNITIVE REAPPRAISAL AND WELL-BEING

Personality Taxonomy

The process of understanding and defining personality as traits occupied theorists for much of the 20th century. One of the earliest theories of personality that assumed behaviour is determined by relatively stable traits was proposed by Eysenck (1947). Eysenck argued personality is genetically determined and can be represented by three broad dimensions: Extroversion, Neuroticism and Psychoticism (Eysenck, 1966). Another pioneering trait theorist was Gordon Allport, who extracted words that represented personality traits from English dictionaries and organised them into a three-level hierarchy encompassing a total of 4500 traits (Allport, 1937). Cattell (1950) built on this work by reducing this list to 16 personality factors that he believed were common in all individuals. A consensus began to emerge in the 1980s when researchers recognised personality as having a hierarchical structure composed of specific traits forming a smaller number of general characteristics (Digman, 1997). This led to a synthesis of several models into a single integrated system (Goldberg, 1993). Structural analysis of the many descriptors eventually revealed five broad factors which showed to be remarkably robust and formed the FFM. These five factors were observed in children and adults (Digman, 1997) and across a variety of languages and cultures (Allik, 2005; McCrae & Costa, 1997). Although many personality taxonomies remain in use, the FFM is recognised as the most robust and frequently used model of personality (e.g., Smith, Sherry, Vidovic, Saklofske, Stoeber, & Benoit, 2018; Thomas, Yalch, Krueger, Wright, Markon, & Hopwood, 2013).

Whilst early models of personality suggested the five factors remain stable over the lifespan, modern models conceptualise personality as a dynamic construct that

develops and interacts with a multitude of factors over time. While personality is inherently influenced by our genetic make-up (Krueger & Johnson, 2008), it also responds to life events, maturation and other processes (Fraley & Roberts, 2005). Researchers found that rank-order stability of most factors increase over the lifespan (Roberts & DelVecchio, 2000), with Conscientiousness and Agreeableness generally increasing and Neuroticism, Openness and Extraversion decreasing over time (Roberts, Walton, & Viechtbauer, 2006; Roberts & Mroczek, 2008). People frequently select environments that reinforce their trait disposition possibly contributing to the observed stability in individuals over the life span (Caspi & Shiner, 2006). Similarly, changes to personality have been linked to adverse events and major shifts in social roles and relationships (Kandler et al., 2010).

Personality and Associations with Mental Health

Depression is characterised by low mood and/or loss of pleasure in most activities (NICE, 2018). Depression severity is determined by the number and severity of symptoms as well as degree of functional impairment. The disorder is associated with a lower quality of life in both those with the condition and their relatives (Saarni et al., 2007). Studying the relationship between personality and depression has long been a subject of clinical interest (Kendler & Myers, 2010). Indeed, most of the extant research has been conducted within the FFM of personality (Widiger & Smith, 2008), as "the organisation of psychopathological tendencies has notable parallels with the organisation of the personality dimensions that underlie those tendencies" (Krueger et

al., 2011, p. 325). Mental health classification systems draw on the dimensional nature of the FFM and utilize it to structure diagnostic criteria of several mental health disorders (Trull & Widiger, 2013). As a result, relationships between personality traits and mental health disorders would be expected.

Historically, research exploring links between personality and depression has largely focused on broad personality dimensions (Rector, Bagby, Huta, & Ayearst, 2012). Kotov and colleagues (2010) found that depression is associated with high Neuroticism (d = 1.33), low Conscientiousness (d = -0.90), and low Extraversion (d = -0.62) in both healthy and patient populations. The relationships between depression and the other two factors - Agreeableness and Openness – however, were found to be weak and inconsistent across studies suggesting the need to focus on associations with lower-order facets (Kotov et al., 2010).

Models of personality based on the FFM split agreeableness into several lower-order facets. These facets reflect specific patterns of thought, emotion, motivation, and behaviour. One of the most widely used personality inventories based on the FFM, the Revised NEO Personality Inventory (NEO-PI-R), divides Agreeableness into Trust, Compliance, Altruism, Straightforwardness, Modesty, and Tender-mindedness (Costa & McCrae, 1992). Other FFM inventories include facets with different labels; for example, the HEXACO inventory separates Agreeableness into Forgivingness, Gentleness, Flexibility, and Patience (Ashton & Lee, 2009). Due to space limitations, this review will primarily explore the relationships between depression and facets of Agreeableness based on the six facets of the NEO-PI-R, as this inventory is most commonly used in the personality literature (Ashton, 2013).

A Theoretical Basis for the relationship between facets of Agreeableness and Depression

Trust. Individuals scoring high on Trust are likely to be forgiving, but may also be naïve. Those scoring low tend to be suspicious and wary of others (Costa & McCrae, 1992). A recent study found that trust enables the development of social capital, which in turn protects from loneliness, lack of support, and consequently, depression (Han et al., 2018). A separate study found low interpersonal trust to be a risk factor for new-onset depression (Kim, Yoon, Kim & Kim, 2017). Based on the evidence, the Trust facet is likely to be negatively associated with depression.

Compliance. Compliance is defined as an individual's response to interpersonal conflict. Those rating high on Compliance are described as meek, docile and cooperative, whilst those who score low tend to be headstrong, antagonistic, and intolerant (Costa & McCrae, 1992). Whilst being compliant with, for example, pharmacological (Goetghebeur & Lapp, 1997) or therapeutic (Edelman & Chambless, 1993) treatments could help outcomes, too much compliance may negatively impact on individual's social rank, which in turn predicts health and wellbeing (Sapolsky, 2005). Consequently, this facet is likely to be both positively and negatively associated with depression or 'balance out' such that the net association with depression is near-zero.

Altruism. Altruistic individuals are described as generous, courteous, and kind, whilst those scoring low on this facet tend to be selfish, greedy, and unwilling (Costa & McCrae, 1992). Although the construct of altruism remains controversial in the literature

with many doubting its validity (Fehr & Fischbacher, 2003; De Waal, 2008), a recent meta-analysis found that performing acts of kindness has a small to medium effect on the well-being of the actor (Curry et al., 2018). This suggests a theoretical basis for Altruism to be negatively associated with depression.

Straightforwardness. Those with high scores on Straightforwardness are characterised as sincere and frank, whereas those with low scores tend to be clever and Machiavellian (Costa & McCrae, 1992). Practising honesty in interpersonal relationships has been shown to improve mental health (Kelly & Wang, 2012). High Straightforwardness could foster the development of honest and genuine relationships, free of deceit and calculation. This could protect the individual at times of need as difficulties need first to be disclosed and acknowledged for healing to begin (Farber, Berano, & Capobianco, 2004). Based on this evidence, the Straightforwardness facet could be theorised to be somewhat negatively related to depression.

Modesty. Individuals scoring high on Modesty can be described as humble and self-effacing, whilst those rating low tend to be self-aggrandizing, haughty, and arrogant (Costa & McCrae, 1992). Modesty has been found to be positively associated with psychological well-being (Aghababaei et al., 2016), although other studies identified Modesty as a risk factor, suggesting self-enhancement protects from depression, although this link was entirely mediated by self-esteem (Sedikides, Rudich, Greggs, Kumashiro, & Rusbult, 2004; Taylor, Lerner, Sherman, Sage, & McDowell, 2003). This conflicting evidence indicates that both positive and negative associations between Modesty and depression could be expected.

Tendermindedness. Tendermindedness is conceptualised as the tendency to determine judgements and attitudes by emotion. Individuals scoring high on this facet are likely to be sympathetic and soft-hearted, while those with low scores tend to be obstinate and logical (Costa & McCrae, 1992). Tendermindedness, labelled as Sympathy in other FFM inventories, has been highlighted as one of the most important components to the development of prosocial disposition (Eisenberg et al., 2002). Failure to develop this competency can lead to the development of mental health disorders (Lee, 2009). Conversely, being sympathetic to one's suffering without appropriate self-care can lead to compassion fatigue and depression (Hegney et al., 2014). Based on this evidence, it would be reasonable to expect Tendermindedness to be both positively and negatively associated with depression.

Clinical Implications

Understanding the links between personality traits and depression has significant clinical implications. While the majority of known risk factors for depression are fixed (e.g., family history, demographics), or predict onset of depression only in the short-term (e.g., adversity), personality traits could improve our understanding of possible vulnerabilities for depression long before its onset (Klein, Kotov, & Bufferd, 2011). Personality screens could help identify at-risk individuals to prevent depression in later life as well as tailor intervention to patient needs to optimize treatment responses (Lahey, 2009; Hayward, Taylor, Smoski, Steffens, & Payne, 2013).

Although Kotov's meta-analysis (2010) found no association between Agreeableness and depressive symptoms, recent studies suggest possible relationships with several of this domain's facets. Mongrain and colleagues (2018) found that interventions increasing individuals' altruistic tendencies reduced depression, particularly if they are low on Agreeableness. Consequently, in the case of Agreeableness, treatments that target particular specific lower-order facets may help to optimise positive outcomes. To better understand which facets have the strongest links with depression, this systematic review aims to answer the question: Which facets of agreeableness are associated with depressive symptoms?

Methods

A systematic review aims to collate and synthesise all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question (Higgins & Green, 2011). It attempts to minimise biases by using explicit and systematic methods and therefore aims to provide reliable findings from which readers may draw conclusions (Oxman & Guyatt, 1993). This review adhered to the Preferred Reporting Items for Systematic Review and Meta-analysis Protocol (PRISMA-P), a widely endorsed tool designed to facilitate the development and reporting of systematic reviews (Moher, Liberati, Tetzlaff, Altman, & PRISMA Group, 2009; Moher et al., 2015).

Eligibility Criteria

The inclusion or exclusion of studies for this review was determined by PECO

(Population, Exposure, Comparator, Outcome) criteria (Table 1). The review included all

Table 1

Inclusion and Exclusion Criteria for Eligibility for Systematic Literature Review

Inclusion criteria	Exclusion criteria
Population	Population
• Adults 18 years of age and over	
Exposure	Exposure
• Standardised measures of personality based on the FFM assessing facets of Agreeableness	
Comparator	Limitations
• Healthy / symptom free controls OR • Normative samples of personality measures	• Languages other than English or Czech
Outcome	Outcome
• Diagnosis of depression, dysphoria or dysthymia based on DSM classification criteria AND/OR • Standardised diagnostic measures of depression or psychometric measures of symptoms of depression	• Single-item measures of depression • Depression ratings given by relatives, spouses or friends

Note. DSM = Diagnostic and Statistical Manual of Mental Disorders; FFM – Five Factor

Model

studies in peer-reviewed journals that investigated relationships between any facets of

Agreeableness (as measured by personality inventories based on the FFM) and

depression in adult populations. Examples of personality measures based on the FFM

included the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992),

the HEXACO model of personality structure (HEXACO; Ashton & Lee, 2009), and the

Big Five Inventory (Goldberg, 1993). Studies that investigated relationships between

facets and depression in response to a specific treatment (e.g. using facets as

predictors of outcome following therapeutic interventions) were not included.

Depression was operationalised as (a) a diagnosis of major depressive disorder (MDD),

dysthymia or dysphoria based on the DSM criteria (American Psychiatric Association,

2013), or (b) symptoms of depression meeting clinical threshold as assessed by

standardised diagnostic or psychometric measures such as the Beck Depression

Inventory-II (BDI-II; Beck et al., 1998), Patient Health Questionnaire (PHQ-9; Kroenke,

2001), or Hamilton Depression Rating Scale (HAM-D; Hamilton, 1960). Studies

investigating differences in facets between groups had to compare depressed or

symptomatic adults with healthy adult controls or normative samples of personality

measures. Designs of eligible studies included prospective, retrospective, longitudinal,

cross-sectional, correlational and between-subject designs.

Information Sources

Relevant studies were identified using computerised sources of multidisciplinary

and subject-specific scholarly literature and research supplied by two platforms: (a) Web

of Science, incorporating Science Citation Index Expanded, Social Sciences Citation

Index, Arts & Humanities Citation Index, Emerging Sources Citation Index, Book

Citation Index, Conference Proceedings Citation Index, and (b) Ovid, incorporating

PsycINFO and MEDLINE. Databases were searched from the starting point of each

database through to 19th February 2019. Grey literature was not searched due to the

high number of records identified by the online database search and time constraints.

Citations of identified papers and reference lists of any related literature reviews (e.g.,

Bagby et al., 2008; Klein et al., 2011; Kotov et al., 2010) were checked to identify any

other potentially relevant articles to include in the review. Articles citing any of the

identified papers were also screened.

Search Strategy

Initial scoping search was utilised to generate key search terms as

recommended by the Cochrane Library guidance (Higgins & Green, 2011). The

Cochrane database was repeatedly checked to ensure the review question had not yet

been systematically investigated. Key words of initially identified papers (Kotov et al.,

2010; Jourdy & Petot, 2017) were screened for additional search terms. Next, electronic

databases were examined to retrieve relevant literature based on the PECO criteria.

Field limitations (human, age 18 years and over, English language) were utilised to

narrow the search. Final search terms for facets of Agreeableness (section 1), FFM

personality inventories (section 2) and depression (Section 3) are listed in Table 2. The

search utilised database-specific truncation (e.g., depress* to cover depression and

COGNITIVE REAPPRAISAL AND WELL-BEING

depressive symptoms) as well as Boolean operators to combine search terms within

each section (OR) and across the three sections (AND). The search was limited to titles,

abstracts, human, adulthood 18+ years, and English language.

Table 2

Search Terms for Databases Screened through OVID and Web of Science

	Section 1 Facets of Agreeableness "OR"	Section 2 FFM Personality Inventories "OR"	Section 3 Depression "OR"
Search Terms	Facet, subdomain, subcomponent, lower-order or subfactor or subscale	Agreeable*, Big Five, Five Factor Model, NEO-FFI, NEO-PI, NEO-PI-R, NEO-PI-3, HEXACO, BHI, Big Five Inventory, BFI, personality inventory, personality trait*	Depress*, dysthym*, dysphor*
Combined Search (Title and Abstract screened)	Section 1 AND	Section 2 AND	Section 3

Study Selection

To identify relevant studies, titles and abstracts of all articles identified through

the database search were initially screened against PECO criteria. Relevant studies

were read in full and again assessed for eligibility. Six randomly selected full-text

studies were reviewed for reliability of inclusion and exclusion decision by an

independent researcher. This step yielded excellent inter-rater reliability (Cohen's κ = 1).

COGNITIVE REAPPRAISAL AND WELL-BEING

Reference sections of all included papers were screened for additional studies that

could have been missed in the search strategy as recommended by NICE (2012)

guidelines for conducting systematic reviews. None met the PECO criteria.

Data Extraction

The Cochrane Collaboration recommends using specific tools for assessing risk

of bias in each included study (Higgins & Green, 2011). Data were extracted and

evaluated using the United States National Institute of Health (NIH) National Heart,

Lung and Blood Institute (NHLBI) Quality Assessment Tool for Observational Cohort

and Cross-Sectional Studies (Appendix A). This assessment tool poses 14 quality-

related questions to which the researcher gives a yes/no answer. An overall quality

rating of Good, Fair or Poor is then allocated (Appendix B). An independent researcher

assessed three randomly selected studies using this tool and comparison of quality

ratings from both researchers yielded almost perfect agreement (Cohen's κ = 0.95).

This represented a difference in opinion on 2 out of 42 quality ratings made. Differences

were discussed until a consensus was reached. The strengths and weaknesses of each

paper that were identified in the process of quality evaluation were considered when

analysing and synthesizing data extracted from the included studies.

Results

The process of searching and screening was based on the PRISMA protocol

(Moher et al., 2009) and is detailed in Figure 1. Nine studies were included in the

COGNITIVE REAPPRAISAL AND WELL-BEING

systematic review. Information relevant to the PECO criteria, findings and quality ratings

for each of the nine studies are detailed in Table 3.


```
IDENTIFICATION

    Records identified via          Records           Records identified via
        MEDLINE and              identified via       backward chaining (i.e.,
      PsycINFO in OVID          Web of Science       citations of nine included
         (n = 378)                 (n = 791)          papers, reference lists of
                                                      related literature reviews)
                                                            (n = 27)

SCREENING

                              Total records
                               (n = 1196)

                         Records screened after         Records excluded
                         duplicates removed               (n = 841)
                              (n = 876)

ELIGIBILITY

                                                    Records excluded
                                                       (n = 26)
                          Full-text articles         Reason for exclusion:
                        assessed for eligibility   Articles did not report on the relationship
                              (n = 35)              between facets of agreeableness and
                                                           depression (n = 21),
                                                    investigated facets of agreeableness as
INCLUDED                                          predictors of specific treatment outcome (n = 2),
                                                   used measures of personality incompatible with
                                                          the Five Factor Model (n = 1),
                          Studies included for      used population under 18 years of age (n = 1),
                              synthesis             or full-text was not available in English (n = 1).
                               (n = 9)
```

Figure 1. Results of search strategy and screening for eligibility

Table 3

Summary and Results of Eligible Studies in Alphabetical Order by Author

Authors	Population	Exposure: Measure of Personality	Comparator	Outcome: Diagnosis or Symptoms of Depression	Results and Conclusion	Evaluation & NIH Quality Assessment Rating
1. Bienvenu et al., (2001)	60 adults with MDD in Baltimore Catchment Area (M_{age}=47.2 years, *SD*=12.7, no. of males = 15).	NEO-PI-R (Costa & McCrae, 1992)	158 healthy adult controls in Baltimore Catchment Area (no MDD or other assessed disorders) Not further specified. Matching across cases not reported.	Depression diagnosed by the Diagnostic Interview Schedule (DIS; Eaton et al., 1997) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN; Wing et al., 1990) at follow-up.	Compared to healthy controls, adults with MDD did not significantly differ in any of the six NEO-PI-R facets of agreeableness: **Trust (A1)** (d = -0.28, CI: -0.58 to 0.02, *p* > 0.05); **Straightforwardness (A2)** (d = -0.07, CI: -0.36 to 0.23, *p* > 0.05); **Altruism (A3)** (d = 0.21, CI: -0.08 to 0.51, *p* > 0.05); **Compliance (A4)** (d = -0.26, CI: -0.56 to 0.03, *p* > 0.05); **Modesty (A5)** (d = 0.22. CI: -0.08 to 0.52, *p* > 0.05); **Tender-mindedness (A6)** (d = 0.14, CI: -0.16 to 0.44, *p* > 0.05)	**Strengths:** Population-based sample**;** good psychometric properties of measures; controlled for alcohol and substance use disorders**;** compared subjects with lifetime (trait) depression with healthy subjects **Limitations:** Effect size not provided; did not match across cases, comparison group poorly defined **Quality Rating: Fair**
2. Bienvenu et al., (2004)	132 adults with MDD and 18 adults with dysthymia in Baltimore Catchment Area (M_{age}=not provided, *SD*=not provided, no. of males/females: not provided for either of the two groups)	NEO-PI-R (Costa & McCrae, 1992)	295 healthy adult controls from NEO-PI-R standardisation sample. Not further specified. Matching across cases not reported.	Depression diagnosed by the Schedules for Clinical Assessment in Neuropsychiatry (SCAN; Wing et al., 1990).	Compared to healthy controls, adults with MDD did not significantly differ in any of the six NEO-PI-R facets of agreeableness (*p* > .002). *Note on analysis.* Adults with dysthymia were not compared to healthy controls due to low sample. Data for tests of difference not provided. Used α = .002 as they corrected for multiple comparisons with the Bonferroni method (α = .05 / number of comparisons [30])	**Strengths:** Population-based sample; MDD diagnosed by a psychiatrist; good psychometric properties of measures **Limitations:** Did not report data on tests of difference; characteristics of MDD group not provided; MDD group comprised of subjects from two different studies; NEO-PI-R standardisation sample not defined; used α = .002 **Quality Rating: Poor**

COGNITIVE REAPPRAISAL AND WELL-BEING

| 3. Hayward et al., (2013) | 112 older psychiatric patients with MDD from Duke Psychiatric Service (M_{age}=69.45 years, SD=6.08, no. of males=43). | NEO-PI-R (Costa & McCrae, 1992) | 104 community-dwelling older adults from central North Carolina with no evidence of depression as indicated by the National Institute of Mental Health Diagnostic Interview Schedule (Robins, Helzer, Croughan, & Ratcliff, 1981) (M_{age}=71.45, SD=5.62, no. of males=31) | Depression diagnosed by Montgomery-ÅAsberg Depression Rating Scale (MADRS; Montgomery & Åsberg, 1979). Depression severity measured at baseline, 3-month follow-up and 12-month follow-up. Depression severity data predated personality assessment. | (a) Scores on NEO-PI-R facets of agreeableness did not make a significant difference in the odds of being diagnosed with depression:
Trust (A1) (OR 0.95, 95% CI: 0.90 to 1.00, p = 0.105);
Straightforwardness (A2) (OR 0.99, CI: 0.95 to 1.04, p = 0.999);
Altruism (A3) (OR 0.99, CI: 0.95 to 1.02, p = 0.995);
Compliance (A4) (OR 0.97, CI: 0.92 to 1.01, p = 0.341);
Modesty (A5) (OR 1.02. CI: 0.98 to 1.07, p = 0.815);
Tender-mindedness (A6) (OR 1.01, CI: 0.96 to 1.05, p = 0.999)

(b) No significant relationships were detected between NEO-PI-R facets of agreeableness and baseline depression severity:
Trust (A1) (β = 0.03, CI: -0.17 to 0.22, p = 0.999);
Straightforwardness (A2) (β = -0.06, CI: -0.29 to 0.17, p = 0.999);
Altruism (A3) (β = 0.05, CI: -0.09 to 0.18, p = 0.995);
Compliance (A4) (β = 0.03, CI: -0.18 to 0.24, p = 0.999);
Modesty (A5) (β = 0.05, CI: -0.16 to 0.25, p = 0.999);
Tender-mindedness (A6) (β = -0.05, CI: -0.23 to 0.14, p = 0.999)

(c) No significant relationships were detected between NEO-PI-R facets of agreeableness and depression severity at 3-month follow-up:
Trust (A1) (β = -0.08, CI: -0.32 to 0.17, p = 0.998);
Straightforwardness (A2) (β = 0.17, CI: -0.13 to 0.46, p = 0.753);
Altruism (A3) (β = -0.06, CI: -0.24 to 0.11, p = 0.993);
Compliance (A4) (β = -0.03, CI: -0.29 to 0.24, p = 0.999);
Modesty (A5) (β = 0.15, CI: 0.10 to 0.41, p = 0.697);
Tender-mindedness (A6) (β = 0.03, CI: -0.21 to 0.26, p = 0.999)

(d) No significant relationships were detected between NEO-PI-R facets of agreeableness and depression severity at 12-month follow-up:
Trust (A1) (β = -0.07, CI: -0.33 to 0.20, p = 0.999);
Straightforwardness (A2) (β = 0.05, CI: -0.27 to | **Strengths:** Study population clearly specified and defined; good psychometric properties of measures; controlled for confounding variables

Limitations: Sample not from general population; MDD and healthy samples recruited through different procedures; personality measured several years after depression severity measurement

Quality Rating: Fair |

					0.36, p = 0.999); **Altruism (A3)** (β = -0.10, CI: -0.29 to 0.08, p = 0.767); **Compliance (A4)** (β = 0.06, CI: -0.22 to 0.35, p = 0.999); **Modesty (A5)** (β = 0.17, CI: -0.11 to 0.44, p = 0.666); **Tender-mindedness (A6)** (β = 0.04, CI: -0.21 to 0.29, p = 0.999) *Note on analysis.* Adjusted for multiple comparisons. Adjusted for sex, age, race, and years of education; personality is treated as a retrospective measure in this study and is used in the analysis of depression severity data collected earlier.	
4. Jourdy, Petot, & Aguerre (2018)	43 patients with depressive disorder from an inpatient & outpatient clinic in Paris (M_{age}=41.79 years, *SD*=11.26, no. of males=23).	NEO-PI-R (Costa & McCrae, 1992)	N/A	Beck Depression Inventory (BDI-II; Beck et al., 1998) administered to depressed patients at Time 1 (M_{score}=29.73, *SD*=12.47) and 12 months later at Time 2 (M_{score}=22.49, *SD*=14.97)	**Modesty (A5)** was positively related to BDI-II scores at Time 2 (β = 0.39, CI: 0.22 to 0.55, p < 0.0001). No other NEO-PI-R facets of agreeableness were associated with depression (statistical data on these facets not provided). *Note on analysis.* The analysis involved entering all personality facets as joint predictors of depression severity at Time 2. Controlled for Time 1 depression.	**Strengths:** Study population clearly specified and defined; good psychometric properties of measures; controlled for Time 1 depression **Limitations**: Low generalizability, low N and high drop-out between Time 1 and Time 2 measurements; possible bias due to subjects receiving various treatment; personality assessed while patients depressed **Quality Rating: Poor**
5. Jourdy & Petot (2017)	58 adults diagnosed with MDD recruited from an inpatient & outpatient clinic in Paris (M_{age}=41.79 years, *SD*=11.26, no. of males=23).	NEO-PI-R (Costa & McCrae, 1992)	French standardisation sample of the NEO-PI-R (Costa & McCrae, 1992) Not further specified, not referenced.	Beck Depression Inventory (BDI-II; Beck et al., 1998) (M_{score}=29.31, *SD*=11.60)	(a) Compared to the French validation sample, depressed adults significantly differed in **Trust (A1)** (p < 0.0016, cannot determine other statistics due to lack of standardisation data)**.** There were no other significant differences in any of the other NEO-PI-R facets of agreeableness between the French validation sample and depressed adults. (b) No significant relationships were detected between NEO-PI-R facets of agreeableness and BDI-II scores: **Trust (A1)** (r = -0.25, CI: -0.48 to 0.01, p > 0.0016); **Straightforwardness (A2)** (r = -0.05, CI: -0.30 to	**Strengths:** Study population clearly specified and defined; good psychometric properties of measures **Limitations:** Low generalizability; low N; personality assessed while patients depressed; possible bias due to subjects receiving various treatment; French

					0.21, *p* > 0.0016); **Altruism (A3)** (r = -0.30, CI: -0.52 to -0.05, *p* > 0.0016); **Compliance (A4)** (r = -0.03, CI: -0.29 to 0.23, *p* > 0.0016); **Modesty (A5) (**r = 0.14, CI: -0.12 to 0.38, *p* > 0.0016); **Tender-mindedness (A6)** (r = -0.14, CI: -0.38 to 0.12, *p* > 0.0016).	standardization sample undefined **Quality Rating: Poor**
6. Naragon-Gainey & Watson (2014)	398 to 598 adult home owners living in the community (precise numbers depended on type of analysis, not specified further) (M$_{age}$=not provided, *SD*=not provided, range 18 to 85 years, no. of fe/males not provided)	NEO-PI-R (Costa & McCrae, 1992) Six-Factor Personality Questionnaire (6FPQ; Jackson, Paunonen, & Tremblay, 2000) In this inventory, scales of Dominance, Even-Tempered, Abasement, and Good-Natured correlated with NEO-PI-R Agreeableness. Jackson Personality Inventory–Revised (JPI-R; Jackson, 1994) In this inventory, scales of Risk-Taking and Empathy	N/A	Centre for Epidemiologic Studies–Depression Scale (CES-D; Radloff, 1977) administered at baseline (M$_{score}$=41.18, *SD*=13.55); and 5 years later at Time 2 (M$_{score}$=44.36, *SD*=12.48)	(a) No significant relationships were detected between baseline depression and three facets of agreeableness: **Modesty (A5)** (r = 0.11, CI: 0.01 to 0.21, *p* > 0.01), **Good-Natured (6PFQ)** (r = -0.15, CI: -0.24 to -0.05, *p* > 0.01), and **Empathy (JPI-R)** (r = -0.15, CI: 0.24 to -0.05, *p* > 0.01). N=402 used for CI calculations. (b) No significant relationships were detected between Time 2 depression and three facets of agreeableness: **Modesty (A5)** (r = 0.07, CI: -0.03 to 0.17, *p* > 0.01), **Good-Natured (6PFQ)** (r = -0.22, CI: -0.31 to -0.12, *p* > 0.01), and **Empathy (JPI-R)** (r = -0.23, CI: -0.32 to -0.14, *p* > 0.01). N=402 used for CI calculations. (c) None of the facets of agreeableness were predictive of Time 2 depression scores: **Modesty (A5)** (β = -0.01, CI: not provided, *p* > 0.01); **Good-Natured (6PFQ)** (β = -0.10, CI: not provided, *p* > 0.01); **Empathy (JPI-R) (**β = -0.03, CI: not provided, *p* > 0.01) *Note on analysis.* Many facets of Agreeableness across several FFM inventories were factor analysed. Three facets from three inventories showed highest factor loading. The analysis involved entering all three identified facets of agreeableness as joint predictors of depression at Time 2. Controlled for Time 1 depression.	**Strengths:** Personality assessed before depression; looked at facets across measures; good psychometric properties of measures **Limitations:** Study population poorly defined; participation rate of eligible persons lower than 50%; personality measures completed at different times over the course of several years; only three facets included in analysis **Quality Rating: Fair**

		correlated with NEO-PI-R Agreeableness.				
7. Quilty et al., (2012)	119 adults with a lifetime diagnosis of MDD (no other demographic details provided for this group)	NEO-PI-R (Costa & McCrae, 1992) Big Five Aspect Scales (BFAS; DeYoung et al., 2007) In this inventory, Agreeableness is comprised of Politeness and Compassion.	N/A	Depression diagnosed by the Structured Clinical Interview for DSM-IV (SCID-I/P; First et al., 1995) Depression severity assessed by Hamilton Depression Rating Scale (HAM-D; Hamilton, 1960) $M_{score}=10.35$, $SD=6.90$)	(a) Depression severity was negatively associated with: **Compassion (BFAS)** (r = -0.13, CI: -0.30 to 0.05, $p < 0.05$), **Trust (A1)** (r = -0.30, CI: -0.46 to -0.13, $p < 0.01$), **Straightforwardness (A2)** (r = -0.12, CI: -0.29 to -.06, $p < 0.05$), **Altruism (A3)** (r = -0.13, CI: -0.29 to 0.06, $p < 0.05$), and **Tender-mindedness (A6)** (r = -0.04, CI: -0.22 to 0.14, p > 0.05). Depression severity was positively associated with: **Modesty (A5)** (r = 0.18, CI: -0.22 to 0.14, $p < 0.01$). No significant relationships were detected between depression severity and: **Compliance (A4)** (r = -0.04, CI: -0.22 to 0.14, $p > 0.05$) and **Politeness (BFAS)** (r = -0.03, CI: -0.21 to 0.15, $p > 0.05$). (b) Depression severity was negatively predicted by: **Trust (A1)** (β = -0.32, CI: not provided, $p < 0.01$) and positively predicted by **Modesty (A5)** (β = 0.19, CI: not provided, $p < 0.01$). None of the other NEO-PI-R or BFAS facets of agreeableness were predictive of depression severity: **Compassion (BFAS)** (β = -0.14, CI: not provided, $p > 0.05$), **Politeness (BFAS)** (β = 0.03, CI: not provided, $p > 0.05$), **Straightforwardness (A2)** (β = -0.07, CI: not provided, $p > 0.05$), **Altruism (A3)** (β = -0.02, CI: not provided, $p > 0.05$), **Compliance (A4)** (β = 0.00, CI: not provided, $p > 0.05$), **Tender-mindedness (A6)** (β = 0.10, CI: not provided, $p > 0.05$). *Note on analysis.* The analysis involved entering all facets of agreeableness from both inventories as joint predictors of depression severity. Controlled for a large number of comparisons.	**Strengths:** Used both clinical interview and a scale to assess depression; good psychometric properties of measures **Limitations:** Study population poorly defined; personality assessed while patients depressed; not controlling for confounding variables **Quality Rating: Poor**

8. Xia et al., (2014)	439 Chinese undergraduate students (M_{age}=21.64 years, SD=1.54, no. of males=222).	Mandarin Chinese version of NEO-PI-R (Costa & McCrae, 1992) Good reliability and validity reported for Chinese sample (Yang, 2010)	N/A	Self-Rating Depression Scale (SDS; Wang & Chi, 1984). SDS mean scores not provided.	None of the NEO-PI-R facets of agreeableness were associated with depression scores: **Trust (A1)** ($β$ = -0.05, CI: not provided, p > 0.05), **Straightforwardness (A2)** ($β$ = -0.05, CI: not provided, p > 0.05), **Altruism (A3)** ($β$ = --0.09, CI: not provided, p > 0.05), **Compliance (A4)** ($β$ = 0.01, CI: not provided, p > 0.05), **Modesty (A5)** ($β$ = 0.03, CI: not provided, p > 0.05), **Tender-mindedness (A6)** ($β$ = -0.03, CI: not provided, p > 0.05). *Note on analysis.* The analysis involved entering all facets of the five factors as joint predictors of depression scores.	**Strengths:** Study population clearly specified and defined; large N; good psychometric properties of measures; controlled for gender and age **Limitations:** Low generalisability; depression scores for study population not reported; **Quality Rating: Fair**
9. Yang et al., (1999)	360 outpatient and inpatient adults with a diagnosis of MDD in 13 Chinese cities (no other demographic details provided for this group)	NEO-PI-R (Costa & McCrae, 1992)	Standardisation population for the NEO-PI-R (Costa & McCrae, 2008). "Western", not further specified.	MDD diagnosed by a psychiatrist based on the Chinese Classification of Mental Disorders (CCMD; Chen, 2002)	Compared to the Western normative sample of the NEO-PI-R, Chinese subjects with depression did not significantly differ in any of the six NEO-PI-R facets of agreeableness (p > 0.05, cannot determine other statistics due to absence of standardisation data).	**Strengths:** Large N from multiple sites **Limitations:** Standardisation sample undefined; compared Chinese sample to Western standardisation sample; poorly defined depressed group **Quality Rating: Poor**

Note: MDD = Major depressive disorder; NEO-PI-R = Revised NEO Personality Inventory; DIS = Diagnostic Interview Schedule; SCAN = Schedules for Clinical Assessment in Neuropsychiatry; *SD* = Standard Deviation; MADRS = Montgomery- ÅAsberg Depression Rating Scale; BDI = Beck Depression Inventory; CES-D = Center for Epidemiologic Studies – Depression Scale; 6FPQ = Six-Factor Personality Questionnaire; JPI-R = Jackson Personality Inventory–Revised; BFAS = Big Five Aspect Scales; DSM = Diagnostic and Statistical Manual; HAM-D = Hamilton Depression Rating Scale; SCID (Structured Clinical Interview for DSM Disorders; SDS = Self-Rating Depression Scale, N/A = not applicable; NIH = National Institute of Health

Summary of Included Studies

The mean age of subjects varied considerably as each study applied different inclusion criteria (range of means 21.6 – 69.5 years). Three studies (study 2, 6 and 7) did not specify demographic characteristics of the samples that were used in facet-depression analyses. Two studies recruited undifferentiated populations who were screened for symptoms of depression (study 6 used a community sample, and study 8 used undergraduate students) while seven studies utilized populations where individuals were diagnosed with depression (study 1, 2, 3, 4, 5, 7 and 9).

All nine studies used the Revised NEO Personality Inventory (Costa & McCrae, 1992) to assess facets of Agreeableness. Studies with non-English speaking populations used culturally adjusted local versions (study 4, 5, 8 and 9). Two studies used additional personality inventories based on the FFM. Study 6 used the Six-Factor Personality Questionnaire (Jackson et al., 2000) and Jackson Personality Inventory–Revised (Jackson, 1994) and study 7 used the Big Five Aspect Scales (DeYoung et al., 2007).

Two studies (1 and 3) explored personality differences between depressed cases and healthy controls. Three studies (2, 5 and 9) investigated these differences using standardisation samples of the NEO-PI-R as control groups. Studies 4, 6, 7 and 8 did not use a comparator. All studies were observational, employing cross-sectional (1, 2, 3, 5, 6, 7, 8 and 9) and/or longitudinal (4, 3 and 6) designs.

Included studies assessed depression using at least one standardised diagnostic tool (1, 2, 3, 7 and 9) or symptom measurement scale (4, 5, 6, 7 and

8). Clinical tools used for diagnosing depression were the Diagnostic Interview

Schedule (Eaton et al., 1997; study 1), the Schedules for Clinical Assessment in

Neuropsychiatry (Wing et al., 1990; studies 1 and 2), the Montgomery-ÅAsberg

Depression Rating Scale (Montgomery & Åsberg, 1979; study 3), the Structured

Clinical Interview for DSM-IV (First et al., 1995; study 7), and the Chinese

Classification of Mental Disorders (Chen, 2002; study 9). Scales for measuring

symptoms of depression used by the included studies were the Beck

Depression Inventory (Beck et al., 1998; studies 4 and 5), the Centre for

Epidemiologic Studies–Depression Scale (Radloff, 1977; study 6), the Hamilton

Depression Rating Scale (Hamilton, 1960; study 7), and the Self-Rating

Depression Scale (Wang & Chi, 1984; study 8).

Critical Appraisal

Trust

Among the studies recruiting undifferentiated populations, study 8 found

no relationship between Trust and depressive symptoms in Chinese

undergraduates, reporting a trivial effect size (Cohen, 1992). Although this

study's sample was large ($n = 439$) and the analysis controlled for other facets,

students' depression severity scores were not reported. Based on the student

population, it could be assumed that relative depression severity was low and

variability of scores was limited and this could have limited the likelihood of

significant associations being found. Adjusting alpha level for a large number of

comparisons also reduced power of the test of significance to detect a potential

effect.

In studies that recruited depressed populations and compared them to healthy controls, study 1 found no significant differences in Trust, reporting a small effect size. The study used two diagnostic tools, controlled for other disorders, alcohol and substance use and compared population-based samples. A possible limitation contributing to this difference being non-significant was the relatively small sample of depressed adults ($n = 60$). No significant differences were also reported by study 2 (effect size not provided), which compared a larger group of depressed adults ($n = 132$) with an unspecified group of healthy controls ($n = 295$) from the NEO-PI-R standardisation sample. In this study, however, the authors used a conservative significance threshold, raising the likelihood of type II error. Study 5 found significant differences in Trust between French depressed adults and an undefined French NEO-PI-R healthy standardisation sample. This difference is highly noteworthy as the authors used a Bonferroni correction, thus lowering the significance threshold. However, generalisability of the findings is limited due to the use of a very small and specific depressed sample ($n = 58$) and a standardisation sample, which may have differed from the depressed sample in important characteristics. The study also did not report statistics that would indicate the effect size of this relationship, although, based on the sample size and reduced alpha level, it could be estimated to be medium to large. Study 9 found no significant differences in Trust between a large sample of depressed Chinese adults ($n = 360$) and a healthy standardisation sample of the NEO-PI-R. The study did not provide any statistical data, which prevents an estimation of effect size. Furthermore, the analysis compared depressed Chinese adults to a healthy Western standardisation sample, which potentially biased the study results.

Study 3 used a depressed group (n = 112) and a healthy community sample (n = 104) and found that scores on Trust made no significant difference in the odds of having a historical diagnosis of depression. This study treated personality as a retrospective measure, therefore temporal ordering of the relationship between trust and depression is not clear and this potentially affects the generalisability of the study's findings. The study also did not report any statistical data for the results of the test of difference. None of the studies tested Trust as a prospective predictor of diagnosis of depression using healthy populations with longitudinal follow-ups.

Several studies also looked at relationships between Trust and depression within depressed populations. In study 7, depression severity was found to be negatively associated with scores on Trust with a small effect size, such that depressed adults with lower scores on Trust had higher severity of depression. Although the authors controlled for other agreeableness facets, the sample was poorly defined and confounding variables were not accounted for. This introduces a potential bias to the study findings. Finally, study 3 examined depressed older adults and found no significant relationship between Trust and baseline depression severity, depression severity at 3-month follow-up, and 12-month follow-up. Effect sizes were trivial.

Overall, no relationship was found between Trust and depressive symptoms in an undifferentiated population, although this evidence is based on only one study. None of the included studies tested this facet as a longitudinal predictor of depression severity in a non-clinical sample. Other findings from the included studies suggest that Trust could potentially be lower in depressed

compared to healthy adults and further decrease in depressed adults with an increase in symptom severity. Nonetheless, the evidence supplied by the two studies is very limited with significant findings not being replicated by other studies, which suggests that the true strength of the relationship is likely to be very weak.

Modesty

In studies recruiting undifferentiated populations, study 8 reported a trivial non-significant relationship between Modesty and depressive symptoms in Chinese undergraduates. The aforementioned strengths and limitations of this study should be considered. Study 6 analysed secondary data from a large community sample and found no relationship between Modesty and baseline depression. Longitudinally, Modesty was also not predictive of depression severity five years after baseline. A large sample size, controlling for other facets and personality being assessed before depression contribute to the strength of this study, although drop-out rate was high (over 50%), potentially confounding the study findings. Reported effect sizes were trivial.

In studies that recruited depressed populations, study 1 found a small, non-significant difference in Modesty between depressed adults and healthy controls, but the sample size was small. No significant differences between depressed and healthy adults were also reported in study 2, although this study used a conservative significance threshold (effect size not provided). Study 5 found a small, non-significant difference in Modesty between French depressed adults and the standardisation sample. Because of the low significance

threshold level determined by the authors using Bonferroni correction, it would have been useful to know the effect size, but the study did not report any statistical data for this relationship. Study 9 found no significant differences in Modesty between Chinese adults and the NEO-PI-R standardisation sample and did not report effect size. Study 3 found that scores on Modesty made no significant differences in the odds of having a historical diagnosis of depression. None of the studies tested Modesty as a prospective predictor of diagnosis of depression using healthy populations with longitudinal follow-ups.

Three studies also looked at associations between Modesty and depression within depressed populations. In study 7, depression severity in depressed adults had a small positive relationship with scores on Modesty. In study 3, scores on Modesty were not related to depression severity at baseline, 3-month follow-up and 12-month follow-up (trivial effect sizes). Longitudinally, in study 4, Modesty in depressed adults had small- to medium-sized positive associations with depression severity after 12 months of treatment. However, similar to study 5 with which it seems to have shared its study sample, study 4 used a small and highly specific sample ($n = 43$) and drop-out rate between baseline and follow-up was high. It is unclear in which direction the high drop-out may have influenced the results.

Overall, two studies reported no relationship between Modesty and depressive symptoms in an undifferentiated population. Longitudinally, one study found Modesty not to predict depression severity five years after baseline. Other findings suggest that depressed individuals who score high on Modesty may experience worse symptoms of depression at the time of testing and 12

months later; however, the evidence is again very limited with significant findings not being replicated, which suggests that the true strength of the relationship is likely to be very weak.

Straightforwardness, Altruism, Tender-mindedness, and Compliance

Only study 8 explored the relationship between symptoms of depression and these four NEO-PI-R facets in an undifferentiated population and found no significant associations with trivial to small effect sizes. None of the studies explored this relationship longitudinally using the four facets as predictors of depression severity in this population.

In studies that recruited depressed populations, study 1, 2, 5 and 9 found no significant differences in these facets between controls and depressed individuals. Again, all effect sizes were trivial to small. None of the studies tested these facets as prospective predictors of diagnosis of depression using healthy populations with longitudinal follow-ups.

Within depressed populations, study 4 did not find a significant relationship between depression severity and the four facets. Study 7 found significant associations between depression severity and three facets: Straightforwardness, Altruism, and Tender-mindedness, but the effect sizes of the relationships were trivial. Scores on Compliance were not related to depression severity in this study. Additionally, some confidence intervals that were calculated using the study data crossed the zero mark (despite authors reporting significant results), which, together with the lack of control for confounding variables and a poorly defined sample, undermines the overall validity of these findings rendering them anomalous. The relationships for the

three facets also became nonsignificant when other facets were controlled for, suggesting none of the three facets were uniquely associated with depressive symptoms. In study 3, scores on Straightforwardness, Altruism, Tender-mindedness, and Compliance, were not related to depression severity at baseline, 3-month and 12-month follow-ups (all trivial effect sizes). In study 4, scores on these facets were found not to be predictive of depression severity after 12 months of treatment when controlling for each other. No data were provided for these results.

Overall, only one study reported relationships between Straightforwardness, Altruism, Tender-mindedness, Compliance and depressive symptoms in an undifferentiated population, which were null. Four studies found no differences in these facets between healthy adults and depressed individuals. None of the studies tested the four facets as prospective predictors of diagnosis of depression using healthy populations with longitudinal follow-ups. Within depressed populations, one out of two studies reported significant associations between Straightforwardness, Altruism, and Tender-mindedness, but methodological issues of the study largely invalidate these findings. Therefore, the evidence accrued from the included studies suggests there is no relationship between these four facets and depression.

Other Facets of Agreeableness

Two studies investigated facets of Agreeableness that were not based on the NEO-PI-R. Study 6 tested two facets for association with depression - Good-Natured and Empathy - of two personality taxonomies (Six-Factor

Personality Questionnaire; 6PFQ, and Jackson Personality Inventory–Revised; JPI-R). These relationships were not significant and effect sizes were trivial.

Study 7 used two facets from the Big Five Aspect Scales (BFAS): Compassion and Politeness. Although the study reported that Politeness was associated with depression severity in depressed adults, the aforementioned methodological limitations of this study together with the trivial effect size question the robustness of this finding. Compassion was not significantly associated with depression severity. Neither facet was associated with depression severity when other facets were controlled for.

Overall, despite the non-significant or unreliable findings reported by the two included studies, this review did not accrue enough data to establish whether relationships between these four non-NEO-PI-R facets and depression exist. Finally, whilst all four facets from the non-NEO-PI-R taxonomies are reported to correlate with Agreeableness, conceptual links with the six NEO-PI-R facets are unclear (e.g., Melchers et al., 2016).

Discussion

The review provided some very limited evidence of an association between depressive symptoms and two NEO-PI-R facets of Agreeableness in depressed adults: Trust and Modesty. Associations between the other four NEO-PI-R facets Straightforwardness, Altruism, Compliance, and Tender-Mindedness, were not demonstrated in any of the included studies. As a result, the review question of which facets of Agreeableness are associated with

depressive symptoms can be tentatively answered for Trust and Modesty, and these findings are discussed in light of past theory and learning outcomes that emerged from this review.

The review methodology had several strengths. A range of literature databases was considered. The database search yielded a large number of records which included studies identified through screening of reference sections of the nine identified studies as well as relevant review papers. This indicates the literature search reached a good level of saturation. Furthermore, risk of bias was assessed using a robust quality assessment tool and the inter-rater reliability of quality ratings between the primary and independent researchers was excellent. Weaknesses of the review methodology include the search being limited to English language and the author not searching grey literature due time constraints. The exclusion of grey literature could present publication bias, making the results of the review more significant than it would be representative. Finally, inter-rater quality assessment of the included studies was only applied to three randomly selected papers.

Findings of this review provide some indication as to the reasons why Kotov et al.'s meta-analysis (2010) did not find a relationship between Agreeableness and depression. It appears the association is complex as some facets may not be linked with depression at all, but others show counteracting associations that balance each other out. Despite Agreeableness being characterised as a domain of interpersonal behaviour and quality of interactions (Costa, McCrae, & Dye, 1991) and depression being known to impact on social life (Steger & Kashdan, 2009), scores on Agreeableness may be of little clinical

value for the treatment of depression. Instead, attention should be paid to the lower order facets which appear to have some associations with the disorder and could be potentially useful in optimising treatment.

While the reviewed evidence does not indicate causality in the reported associations, theory indicating possible causal links should be considered. Several theoretical underpinnings could explain the negative relationship between depression and Trust. For instance, Beck's cognitive theory (1967) posited that depressed individuals demonstrate negative automatic thinking whereby critical thoughts about self, the world and the future occur spontaneously, and this reinforces faulty information processing creating further cognitive bias. For example, depressed individuals may draw negative conclusions in the absence of supporting data and focus on the worst aspects of a situation (Beck, 1967). This bias may not only influence self-to-self relating, but also impact on the quality of relationships with others through avoidance of intimacy or fear of losing others. Low interpersonal trust has been found to be linked with low social capital, loneliness, and lack of support (Han et al., 2018). It is possible that, in depression, trust in self and others may diminish with confidence and subsequent avoidance of social situations. This results in lack of positive reinforcement and non-satisfaction of need for relatedness (Ryan & Deci, 2000).

There are also possible clinical implications of the potential relationship between depression and Trust. For instance, having trust in the therapist and treatment has been identified as one of the key underlying factors in psychotherapy research (e.g., Leach, 2005; Marshall & Serran, 2004). Focusing

on developing a trusting therapeutic relationship early in the treatment process may therefore be particularly beneficial when working with depressed individuals.

The positive relationship between Modesty and depressive symptoms may appear counter-intuitive as this personality trait is conventionally perceived as a virtue valued by others (McMullin, 2010). Nonetheless, modesty seems to have its pitfalls, as those exhibiting this trait may be more likely to deny their need for support when unwell (McMullin, 2010). Therefore, it may be the case that Modesty ultimately hinders acquisition of resources at the time of need and leads to higher vulnerability. In addition, being modest, particularly in men, can be perceived by others as a sign of weakness and low social status (Moss-Racusin, Phelan, & Rudman, 2010). This may lead to subordination, social defeat and negative self-image, well-documented factors playing role in depression (Gilbert, 2006). In light of the reviewed literature, the potential relationship between depression and modesty could have important clinical implications. High modesty in depressed individuals could be identified as a potential barrier to accessing support. Therefore, it may be particularly useful for clinicians to discuss modesty early in the treatment process to empower depressed individuals to acknowledge their needs and engage in treatment in order to meet them. Potential consequences of high modesty could be explored with depressed clients and incorporated in formulation to inform treatment.

Theoretical evidence supporting possible associations between the four remaining NEO-PI-R facets and depression in either direction exists and this, together with the low quality of evidence retrieved from the included papers,

perhaps warrants further experimental studies focusing on these constructs.

Furthermore, it may be the case that relationships between some personality

facets and depression are non-linear, and this would not have been identified by

the included studies.

For instance, whilst average to high forms of altruism may be beneficial

for well-being (Curry et al., 2018), it may be harmful at its extreme. This is well

demonstrated by the theory of unmitigated communion, where individuals who

involve themselves with others to the exclusion of themselves neglect their core

needs leading to lower well-being and more depressive symptoms (Helgeson &

Fritz, 1998). Similarly, compliance - whilst perhaps benign in its milder form,

high levels are linked with low social rank, predicting poorer health (Sapolsky,

2005). The indication of an association between straightforwardness and

depression is less clear. Whilst generally seen as a virtue affording honest and

genuine relationships (Faber et al., 2004), individuals scoring very high on this

personality trait could be perceived as blunt, insensitive or coarse and have

lower social capital. Finally, tendermindedness also appears to have a

conflicting relationship with depression, with average levels being possibly

benign and extreme levels likely to cause compassion fatigue. Studies aiming to

explore associations between the four facets and depression could test

predictions about non-linear relationships.

Limitations

This review had several limitations. While symptoms or diagnosis of

depression were generally among the primary variables of interest across the

included studies, facets of Agreeableness and their relationships with depression were often reported as an adjunct to other, more detailed results of analyses authors had a primary interest in. Consequently, some of the research designs employed in the included studies are arguably poorly suited to answer the review question.

Related to this is the overall poor quality of retrieved evidence that informs the review question. Studies often used small and specific samples or did not provide sample characteristics, statistical data was missing and drop-outs were high or unreported. In addition, whilst some indication of a relationship between depression and Trust and Modesty was found, evidence comes from only three studies (two of which appear to have shared their sample) and variance accounted for by the facets was only very small. Confidence in these relationships is therefore low. The review also did not accrue enough evidence to sufficiently explore relationships between depression and other, non-NEO-PI-R facets of Agreeableness.

Finally, it is possible that cognitive biases that are symptomatic of depression impact on the self-report scores on personality facets such that the scores themselves are more representative of depression than premorbid personality traits. In other words, individuals' self-report of their personality could differ between depressed and non-depressed. Perhaps this could be mitigated by collecting other-report data. The fact that the included studies did not explore links between the personality of healthy adults and the development of depression in later life highlights an important limitation of this review.

Future Research

Although the results of the current review indicate that Trust and Modesty are associated with depression, this would preferably be confirmed by conducting a more methodologically robust empirical study. This study could employ a longitudinal, prospective design using healthy populations to explore links between pre-morbid personality facets and incidence of depression in later life. Assuming key covariates such as age, sex or marital status (Meng et al., 2017) would be controlled for, this design could offer findings with high clinical significance.

Alternatively, links between facets and depression could also be assessed through personality states. These are short-term, concrete patterns of acting, feeling, and thinking (Heller, Komar, & Lee, 2007) and possible study designs could explore whether depressive symptoms covary with states and behaviours that are high in certain facets of agreeableness.

Conclusions

The systematic review provides some limited evidence indicating NEO-PI-R facets Trust and Modesty are associated with depression. Specifically, it appears that in depressed adults Trust decreases and Modesty increases with depression severity. Whilst causality of these links cannot be established, considering these associations may be of clinical value when working with depressed individuals. Despite conceptual evidence suggesting that links between depression and other facets could exist, none were found to reach

significance. Future research could (a) focus on further empirical testing of the reported associations through more robust longitudinal designs, and (b) employ methods of analysis that could uncover potential non-linear associations between the other four facets.

References

Aghababaei, N., Błachnio, A., Arji, A., Chiniforoushan, M., Tekke, M., &

Mehrabadi, A. F. (2016). Honesty–humility and the HEXACO structure of

religiosity and well-being. *Current Psychology*, *35*, 421-426.

doi:10.1007/s12144-015-9310-5

Allik, J. (2005). Personality dimensions across cultures. *Journal of Personality

Disorders*, *19*, 212-232. doi:10.1521/pedi.2005.19.3.212

Allport, G. W. (1937). *Personality: a psychological interpretation*. Oxford,

England: Holt. doi:10.2307/2785590

American Psychiatric Association. (2013). *Diagnostic and statistical manual of

mental disorders (5th ed.)*. Washington, DC: Author. doi:10.1108/rr-10-

2013-0256

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the

major dimensions of personality. *Journal of Personality Assessment*, *91*,

340-345. doi:10.1080/00223890902935878

Ashton, M. C. (2013). *Individual differences and personality*. Academic Press.

Bienvenu, O. J., Nestadt, G., Samuels, J. F., Costa, P. T., Howard, W. T., &

Eaton, W. W. (2001). Phobic, panic, and major depressive disorders and

the five-factor model of personality. *The Journal of Nervous and Mental

Disease*, *189*, 154-161. doi:10.1097/00005053-200103000-00003

Bienvenu, O. J., Samuels, J. F., Costa, P. T., Reti, I. M., Eaton, W. W., & Nestadt, G. (2004). Anxiety and depressive disorders and the five-factor model of personality: A higher-and lower-order personality trait investigation in a community sample. *Depression and Anxiety*, *20*, 92-97. doi:10.1002/da.20026

Caspi, A., & Shiner, R. L. (2006). Personality Development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 300-365). Hoboken, NJ, US: John Wiley & Sons Inc.

Cattell, R. B. (1950). *Personality: A systematic theoretical and factual study* (1st ed.). New York, NY, US: McGraw-Hill. doi:10.1037/10773-000

Chen, Y. F. (2002). Chinese classification of mental disorders (CCMD-3): towards integration in international classification. *Psychopathology*, *35*, 171-175. doi:10.1159/000065140

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159. doi:10.1037//0033-2909.112.1.155

Costa Jr, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and individual Differences*, *12*, 887-898. doi:10.1016/0191-8869(91)90177-d

Costa Jr, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, *6,* 343-359. doi:10.1521/pedi.1992.6.4.343

Costa, P. T., & McCrae, R. R. (2008). The revised neo personality inventory (neo-pi-r). *The SAGE Handbook of Personality Theory and Assessment*, *2*, 179-198. doi:10.4135/9781849200479.n9

Curry, O. S., Rowland, L. A., Van Lissa, C. J., Zlotowitz, S., McAlaney, J., & Whitehouse, H. (2018). Happy to help? A systematic review and meta-analysis of the effects of performing acts of kindness on the well-being of the actor. *Journal of Experimental Social Psychology*, *76*, 320-329. doi:10.1016/j.jesp.2018.02.014

De Waal, F. B. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology*, *59*, 279-300. doi:10.1146/annurev.psych.59.103006.093625

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *73*, 1246-1256. doi:10.1037//0022-3514.73.6.1246

Eaton, W. W., Anthony, J. C., Gallo, J., Cai, G., Tien, A., Romanoski, A., ... & Chen, L. S. (1997). Natural history of Diagnostic Interview Schedule/DSM-IV major depression: The Baltimore epidemiologic catchment area follow-up. *Archives of General Psychiatry*, *54*, 993-999. doi:10.1001/archpsyc.1997.01830230023003

Edelman, R. E., & Chambless, D. L. (1993). Compliance during sessions and

homework in exposure-based treatment of agoraphobia. *Behaviour

Research and Therapy*, *31*, 767-773. doi:10.1016/0005-7967(93)90007-h

Eysenck, H. J. (1947). Primary social attitudes. *International Journal of Opinion

and Attitude Research*, *1*, 49-84. doi:10.1037/h0054347

Eysenck, H. J. (1966). Personality and experimental psychology. *Bulletin of the

British Psychological Society*, *19*, 1-28. doi:10.1037/0022-

3514.73.6.1224

Farber, B. A., Berano, K. C., & Capobianco, J. A. (2004). Clients' Perceptions of

the Process and Consequences of Self-Disclosure in Psychotherapy.

Journal of Counseling Psychology, *51*, 340-346. doi:10.1037/0022-

0167.51.3.340

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*,

785-791. doi:10.1038/nature02043

Fraley, R. C., & Roberts, B. W. (2005). Patterns of continuity: a dynamic model

for conceptualizing the stability of individual differences in psychological

constructs across the life course. *Psychological Review*, *112*, 60-74.

doi:10.1037/0033-295x.112.1.60

Gilbert, P. (2006). Evolution and depression: issues and implications.

Psychological Medicine, *36*, 287-297. doi:10.1017/s0033291705006112

Goetghebeur, E., & Lapp, K. (1997). The effect of treatment compliance in a

placebo-controlled trial: regression with unpaired data. *Journal of the*

Royal Statistical Society: Series C (Applied Statistics), *46*, 351-364.

doi:10.1111/1467-9876.00074

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American*

Psychologist, *48*, 26-34. doi:10.1037//0003-066x.48.1.26

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neuro-*

surgery, and Psychiatry. 23, 56–61. doi:10.1136/jnnp.23.1.56

Han, K. M., Han, C., Shin, C., Jee, H. J., An, H., Yoon, H. K., ... & Kim, S. H.

(2018). Social capital, socioeconomic status, and depression in

community-living elderly. *Journal of Psychiatric Research*, *98*, 133-140.

doi:10.1016/j.jpsychires.2018.01.002

Hayward, R. D., Taylor, W. D., Smoski, M. J., Steffens, D. C., & Payne, M. E.

(2013). Association of five-factor model personality domains and facets

with presence, onset, and treatment outcomes of major depression in

older adults. *The American Journal of Geriatric Psychiatry*, *21*, 88-96.

doi:10.1016/j.jagp.2012.11.012

Hegney, D. G., Craigie, M., Hemsworth, D., Osseiran-Moisson, R., Aoun, S.,

Francis, K., & Drury, V. (2014). Compassion satisfaction, compassion

fatigue, anxiety, depression and stress in registered nurses in Australia:

study 1 results. *Journal of Nursing Management*, *22*, 506-518.

doi:10.1111/jonm.12160

Helgeson, V. S., & Fritz, H. L. (1998). A theory of unmitigated communion. *Personality and Social Psychology Review*, *2*, 173-183. doi:10.1207/s15327957pspr0203_2

Heller, D., Komar, J., & Lee, W. B. (2007). The dynamics of personality states, goals, and well-being. *Personality and Social Psychology Bulletin*, *33*, 898-910. doi:10.1177/0146167207301010

Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration.* Retrieved from http://handbook.cochrane.org.

Jackson, D. N. (1994). *Jackson Personality Inventory–Revised manual*. Port Huron, MI: Sigma Assessment Systems. doi:10.4135/9781412952644.n232

Jackson, D. N., Paunonen, S. V., & Tremblay, P. F. (2000). *Six factor personality questionnaire manual*. Port Huron, MI: Sigma Assessment Systems.

Jourdy, R., & Petot, J. M. (2017). Relationships between personality traits and depression in the light of the "Big Five" and their different facets. *L'Évolution Psychiatrique*, *82*, e27-e37. doi:10.1016/j.evopsy.2017.08.002

Jourdy, R., Petot, J. M., & Aguerre, C. (2018). The relationships between personality traits and one-year outcomes in patients with major

depressive disorder without psychotic features. *Revue Européenne de Psychologie Appliquée, 68*, 151-159. doi:10.1016/j.erap.2018.08.003

Kandler, C., Bleidorn, W., Riemann, R., Spinath, F. M., Thiel, W., & Angleitner, A. (2010). Sources of cumulative continuity in personality: A longitudinal multiple-rater twin study. *Journal of Personality and Social Psychology, 98*, 995-1008. doi:10.1037/a0019558

Kelly, A. E., & Wang, L. (2012). *A Life without Lies: Can Living More Honestly Improve Health?* [PowerPoint slides]. Retrieved from https://cbsphilly.files.wordpress.com/2012/08/kelly-a-life-without-lies.pdf

Kendler, K. S., & Myers, J. (2010). The genetic and environmental relationship between major depression and the five-factor model of personality. *Psychological Medicine, 40*, 801-806. doi:10.1017/s0033291709991140

Kim, J. Y., Yoon, J., Kim, M. H., & Kim, S. S. (2017). Association between interpersonal trust, reciprocity, and suicidal behaviors: A longitudinal cohort study in South Korea. *Preventive Medicine, 99*, 218-221. doi:10.1016/j.ypmed.2017.02.023

Klein, D. N., Kotov, R., & Bufferd, S. J. (2011). Personality and depression: explanatory models and review of the evidence. *Annual Review of Clinical Psychology, 7*, 269-295. doi:10.1146/annurev-clinpsy-032210-104540

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: a

meta-analysis. *Psychological Bulletin*, *136*, 768-821.

doi:10.1037/a0020327

Krueger, R. F., & Eaton, N. R. (2010). Personality traits and the classification of

mental disorders: Toward a more complete integration in DSM–5 and an

empirical model of psychopathology. *Personality Disorders: Theory,*

Research, and Treatment, *1*, 97-118. doi:10.1037/a0018990

Krueger, R. F., Eaton, N. R., Derringer, J., Markon, K. E., Watson, D., & Skodol,

A. E. (2011). Personality in DSM–5: Helping delineate personality

disorder content and framing the metastructure. *Journal of Personality*

Assessment, *93*, 325-331. doi: 10.1080/00223891.2011.577478

Lahey, B. B. (2009). Public health significance of neuroticism. *American*

Psychologist, *64*, 241-256. doi:10.1037/a0015309

Leach, M. J. (2005). Rapport: A key to treatment success. Complementary

Therapies in Clinical Practice, *11*, 262-265.

Doi:10.1016/j.ctcp.2005.05.005

Lee, S. A. (2009). Measuring individual differences in trait sympathy: Instrument

construction and validation. *Journal of Personality Assessment*, *91*, 568-

583. doi:10.1080/00223890903228620

Marshall, W. L., & Serran, G. A. (2004). The role of the therapist in offender

treatment. *Psychology, Crime & Law*, *10*, 309-320.

doi:10.1080/10683160410001662799

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*, 509-516. doi:10.1037//0003-066x.52.5.509

McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81-90. doi:10.1037//0022-3514.52.1.81

McCrae, R. R., & Costa, P. T. (2013). Introduction to the empirical and theoretical status of the five-factor model of personality traits. In T. A. Widiger & P. T. Costa (Eds.), *Personality disorders and the five-factor model of personality* (3rd ed., pp. 15–27). Washington, DC: American Psychological Association.

Melchers, M. C., Li, M., Haas, B. W., Reuter, M., Bischoff, L., …& Montag, C. (2016). Similar personality patterns are associated with empathy in four different countries. *Frontiers in Psychology*, *7*, 290. doi:10.3389/fpsyg.2016.00290

Meng, X., Brunet, A., Turecki, G., Liu, A., D'Arcy, C., & Caron, J. (2017). Risk factor modifications and depression incidence: a 4-year longitudinal Canadian cohort of the Montreal Catchment Area Study. *BMJ open*, *7*, e015156. doi:10.1136/bmjopen-2016-015156

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*, 1-6. doi:10.1371/journal.pmed.1000097

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*, 1. doi:10.1186/2046-4053-4-1

Montgomery, S. A., & Åsberg, M. A. R. I. E. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry*, *134*, 382-389. doi:10.1192/bjp.134.4.382

Moss-Racusin, C. A., Phelan, J. E., & Rudman, L. A. (2010). When men break the gender rules: Status incongruity and backlash against modest men. *Psychology of Men & Masculinity*, *11*, 140-151. doi:10.1037/a0018093

Naragon-Gainey, K. & Watson, D. (2014). Consensually defined facets of personality as prospective predictors of change in depression symptoms. *Assessment*, *21*, 387-403. doi:10.1177/1073191114528030

National Institute for Health and Care Excellence. (2012). *The guidelines manual.* Retrieved from http://www.nice.org

National Institute for Health and Care Excellence. (2018). *Depression in adults: recognition and management.* Retrieved from https://www.nice.org.uk/guidance/cg90/chapter/Introduction

Oxman, A. D., & Guyatt, G. H. (1993). The Science of Reviewing Researcha. *Annals of the New York Academy of Sciences*, *703*, 125-134. doi:10.1111/j.1749-6632.1993.tb26342.x

Quilty, L. C., Pelletier, M., DeYoung, C. G., & Bagby, M. R. (2013). Hierarchical personality traits and the distinction between unipolar and bipolar disorders. *Journal of Affective Disorders*, *157*, 247-254. doi:10.1016/j.jad.2012.11.012

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385-401. doi:10.1177/014662167700100306

Rector, N. A., Bagby, R. M., Huta, V., & Ayearst, L. E. (2012). Examination of the trait facets of the five-factor model in discriminating specific mood and anxiety disorders. *Psychiatry Research*, *199*, 131-139. doi:10.1016/j.psychres.2012.04.027

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3-25. doi:10.1037//0033-2909.126.1.3

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*, 313-345. doi:10.1111/j.1745-6916.2007.00047.x

Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science*, *17*, 31-35. doi:10.1111/j.1467-8721.2008.00543.x

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*, 1-25. doi:10.1037/0033-2909.132.1.1

Robins, L. N., Helzer, J. E., Croughan, J., & Ratcliff, K. S. (1981). National Institute of Mental Health diagnostic interview schedule: Its history, characteristics, and validity. *Archives of General Psychiatry*, *38*, 381-389. doi:10.1001/archpsyc.1981.01780290015001

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54-67. doi:10.1006/ceps.1999.1020

Saarni, S. I., Suvisaari, J., Sintonen, H., Pirkola, S., Koskinen, S., Aromaa, A., & Lönnqvist, J. (2007). Impact of psychiatric disorders on health-related quality of life: general population survey. *The British Journal of Psychiatry*, *190*, 326-332. doi:10.1192/bjp.bp.106.025106

Sapolsky, R. M. (2005). The influence of social hierarchy on primate health. *Science*, *308*, 648-652. doi:10.1126/science.1106477

Sedikides, C., Rudich, E. A., Gregg, A. P., Kumashiro, M., & Rusbult, C. (2004). Are normal narcissists psychologically healthy?: Self-esteem matters. *Journal of Personality and Social Psychology*, *87*, 400–416. doi:10.1037/0022-3514.87.3.400

Steger, M. F., & Kashdan, T. B. (2009). Depression and everyday social activity, belonging, and well-being. *Journal of Counseling Psychology, 56*, 289-300. doi:10.1037/a0015416

Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M., & McDowell, N. K. (2003). Are self-enhancing cognitions associated with healthy or unhealthy biological profiles?. *Journal of Personality and Social Psychology*, *85*, 605-615. doi:10.1037/0022-3514.85.4.605

Thomas, K. M., Yalch, M. M., Krueger, R. F., Wright, A. G., Markon, K. E., & Hopwood, C. J. (2013). The convergent structure of DSM-5 personality trait facets and five-factor model trait domains. *Assessment*, *20*, 308-311. doi:10.1177/1073191112457589

Trull, T. J., & Widiger, T. A. (2013). Dimensional models of personality: the five-factor model and the DSM-5. Dialogues in Clinical Neuroscience, *15*, 135-146. doi:10.1037/10140-003

Smith, M. M., Sherry, S. B., Vidovic, V., Saklofske, D. H., Stoeber, J., & Benoit, A. (2018). Perfectionism and the five-factor model of personality: A meta-analytic review. *Personality and Social Psychology Review*, *1088868318814973*. doi:10.1177/1088868318814973

Wang, Z. Y., & Chi, Y. F. (1984). Self-rating depression scale (SDS). *Shanghai Archives of Psychiatry, 2*, 71-72. doi:10.1001/archpsyc.1965.01730060026004

Wing, J. K., Babor, T., Brugha, T. S., Burke, J., Cooper, J. E., Giel, R., ... &
     Sartorius, N. (1990). SCAN: schedules for clinical assessment in
     neuropsychiatry. *Archives of General Psychiatry, 47*, 589-593.
     doi:10.1001/archpsyc.1990.01810180089012

World Health Organisation. (2018). *Depression.* Retrieved from
     https://www.who.int/en/news-room/fact-sheets/detail/depression

Xia, L. X., Xu, X. Y., Hollon, S. D., & Zhang, J. F. (2014). The relation of self-
     supporting personality, big five personality and depression. *Current
     Psychology*, *33*, 630-643. doi:10.1007/s12144-014-9233-6

Yang, J. F. (2010). Cross-cultural personality assessment: the revised neo
     personality inventory in China. Social Behavior and Personality: *An
     International Journal*, *38*, 1097-1104. doi:10.2224/sbp.2010.38.8.1097

COGNITIVE REAPPRAISAL AND WELL-BEING

Appendix A

NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies

12/11/2017
Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies - NHLBI, NIH

NIH National Heart, Lung, and Blood Institute

Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies

Criteria	Yes	No	Other (CD, NR, NA)*
1. Was the research question or objective in this paper clearly stated?			
2. Was the study population clearly specified and defined?			
3. Was the participation rate of eligible persons at least 50%?			
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?			
5. Was a sample size justification, power description, or variance and effect estimates provided?			
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?			
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?			
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?			
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
10. Was the exposure(s) assessed more than once over time?			
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
12. Were the outcome assessors blinded to the exposure status of participants?			
13. Was loss to follow-up after baseline 20% or less?			
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?			

Quality Rating (Good, Fair, or Poor) (see guidance)
Rater #1 initials:
Rater #2 initials:
Additional Comments (If POOR, please state why):

*CD, cannot determine; NA, not applicable; NR, not reported

Guidance for Assessing the Quality of Observational Cohort and Cross-Sectional Studies

The guidance document below is organized by question number from the tool for quality assessment of observational cohort and cross-sectional studies.

Question 1. Research question

Did the authors describe their goal in conducting this research? Is it easy to understand what they were looking to find? This issue is important for any scientific paper of any type. Higher quality scientific research explicitly defines a research question.

Questions 2 and 3. Study population

Did the authors describe the group of people from which the study participants were selected or recruited, using demographics, location, and time period? If you were to conduct this study again, would you know who to recruit, from where, and from what time period? Is the cohort population free of the outcomes of interest at the time they were recruited?

An example would be men over 40 years old with type 2 diabetes who began seeking medical care at Phoenix Good Samaritan Hospital between January 1, 1990 and December 31, 1994. In this example, the population is clearly described as: (1) who (men over 40 years old with type 2 diabetes); (2) where (Phoenix Good Samaritan Hospital); and (3) when (between January 1, 1990 and December 31, 1994). Another example is women ages 34 to 59 years of age in 1980 who were in the nursing profession and had no known coronary disease, stroke, cancer, hypercholesterolemia, or diabetes, and were recruited from the 11 most populous States, with contact information obtained from State nursing boards.

In cohort studies, it is crucial that the population at baseline is free of the outcome of interest. For example, the nurses' population above would be an appropriate group in which to study incident coronary disease. This information is usually found either in descriptions of population recruitment, definitions of variables, or inclusion/exclusion criteria.

You may need to look at prior papers on methods in order to make the assessment for this question. Those papers are usually in the reference list.

If fewer than 50% of eligible persons participated in the study, then there is concern that the study population does not adequately represent the target population. This increases the risk of bias.

Question 4. Groups recruited from the same population and uniform eligibility criteria

Were the inclusion and exclusion criteria developed prior to recruitment or selection of the study population? Were the same underlying criteria used for all of the subjects involved? This issue is related to the description of the study population, above, and you may find the information for both of these questions in the same section of the paper.

Most cohort studies begin with the selection of the cohort; participants in this cohort are then measured or evaluated to determine their exposure status. However, some cohort studies may recruit or select exposed participants in a different time or place than unexposed participants, especially retrospective cohort studies– which is when data are obtained from the past (retrospectively), but the analysis examines exposures prior to outcomes. For example, one research question could be whether diabetic men with clinical depression are at higher risk for cardiovascular disease than those without clinical depression. So, diabetic men with depression might be selected from a mental health clinic, while diabetic men without depression might be selected from an internal medicine or endocrinology clinic. This study recruits groups from different clinic populations, so this example would get a "no."

However, the women nurses described in the question above were selected based on the same inclusion/exclusion criteria, so that example would get a "yes."

Question 5. Sample size justification

Did the authors present their reasons for selecting or recruiting the number of people included or analyzed? Do they note or discuss the statistical power of the study? This question is about whether or not the study had enough participants to detect an association if one truly existed.

A paragraph in the methods section of the article may explain the sample size needed to detect a hypothesized difference in outcomes. You may also find a discussion of power in the discussion section (such as the study had 85 percent power to detect a 20 percent increase in the rate of an outcome of interest, with a 2-sided alpha of 0.05). Sometimes estimates of variance and/or estimates of effect size are given, instead of sample size calculations. In any of these cases, the answer would be "yes."

However, observational cohort studies often do not report anything about power or sample sizes because the analyses are exploratory in nature. In this case, the answer would be "no." This is not a "fatal flaw." It just may indicate that attention was not paid to whether the study was sufficiently sized to answer a prespecified question–i.e., it may have been an exploratory, hypothesis-generating study.

Question 6. Exposure assessed prior to outcome measurement

This question is important because, in order to determine whether an exposure causes an outcome, the exposure must come before the outcome.

For some prospective cohort studies, the investigator enrolls the cohort and then determines the exposure status of various members of the cohort (large epidemiological studies like Framingham used this approach). However, for other cohort studies, the cohort is selected based on its exposure status, as in the example above of depressed diabetic men (the exposure being depression). Other examples include a cohort identified by its exposure to fluoridated drinking water and then compared to a cohort living in an area without fluoridated water, or a cohort of military personnel exposed to combat in the Gulf War compared to a cohort of military personnel not deployed in a combat zone.

With either of these types of cohort studies, the cohort is followed forward in time (i.e., prospectively) to assess the outcomes that occurred in the exposed members compared to nonexposed members of the cohort. Therefore, you begin the study in the present by looking at groups that were exposed (or not) to some biological or behavioral factor, intervention, etc., and then you follow them forward in time to examine outcomes. If a cohort study is conducted properly, the answer to this question should be "yes," since the exposure status of members of the cohort was determined at the beginning of the study before the outcomes occurred.

For retrospective cohort studies, the same principal applies. The difference is that, rather than identifying a cohort in the present and following them forward in time, the investigators go back in time (i.e., retrospectively) and select a cohort based on their exposure status in the past and then follow them forward to assess the outcomes that occurred in the exposed and nonexposed cohort members. Because in retrospective cohort studies the exposure and outcomes may have already occurred (it depends on how long they follow the cohort), it is important to make sure that the exposure preceded the outcome.

Sometimes cross-sectional studies are conducted (or cross-sectional analyses of cohort-study data), where the exposures and outcomes are measured during the same timeframe. As a result, cross-sectional analyses provide weaker evidence than regular cohort studies regarding a potential causal relationship between exposures and outcomes. For cross-sectional analyses, the answer to Question 6 should be "no."

Question 7. Sufficient timeframe to see an effect

Did the study allow enough time for a sufficient number of outcomes to occur or be observed, or enough time for an exposure to have a biological effect on an outcome? In the examples given above, if clinical depression has a biological effect on increasing risk for CVD, such an effect may take years. In the other example, if higher dietary sodium increases BP, a short timeframe may be sufficient to assess its association with BP, but a longer timeframe would be needed to examine its association with heart attacks.

The issue of timeframe is important to enable meaningful analysis of the relationships between exposures and outcomes to be conducted. This often requires at least several years, especially when looking at health outcomes, but it depends on the research question and outcomes being examined.

Cross-sectional analyses allow no time to see an effect, since the exposures and outcomes are assessed at the same time, so those would get a "no" response.

Question 8. Different levels of the exposure of interest

If the exposure can be defined as a range (examples: drug dosage, amount of physical activity, amount of sodium consumed), were multiple categories of that exposure assessed? (for example, for drugs: not on the medication, on a low dose, medium dose, high dose; for dietary sodium, higher than average U.S. consumption, lower than recommended consumption, between the two). Sometimes discrete categories of exposure are not used, but instead exposures are measured as continuous variables (for example, mg/day of dietary sodium or BP values).

In any case, studying different levels of exposure (where possible) enables investigators to assess trends or dose-response relationships between exposures and outcomes–e.g., the higher the exposure, the greater the rate of the health outcome. The presence of trends or dose-response relationships lends credibility to the hypothesis of causality between exposure and outcome.

For some exposures, however, this question may not be applicable (e.g., the exposure may be a dichotomous variable like living in a rural setting versus an urban setting, or vaccinated/not vaccinated with a one-time vaccine). If there are only two possible exposures (yes/no), then this question should be given an "NA," and it should not count negatively towards the quality rating.

Question 9. Exposure measures and assessment

Were the exposure measures defined in detail? Were the tools or methods used to measure exposure accurate and reliable–for example, have they been validated or are they objective? This issue is important as it influences confidence in the reported exposures. When exposures are measured with less accuracy or validity, it is

harder to see an association between exposure and outcome even if one exists. Also as important is whether the exposures were assessed in the same manner within groups and between groups; if not, bias may result.

For example, retrospective self-report of dietary salt intake is not as valid and reliable as prospectively using a standardized dietary log plus testing participants' urine for sodium content. Another example is measurement of BP, where there may be quite a difference between usual care, where clinicians measure BP however it is done in their practice setting (which can vary considerably), and use of trained BP assessors using standardized equipment (e.g., the same BP device which has been tested and calibrated) and a standardized protocol (e.g., patient is seated for 5 minutes with feet flat on the floor, BP is taken twice in each arm, and all four measurements are averaged). In each of these cases, the former would get a "no" and the latter a "yes."

Here is a final example that illustrates the point about why it is important to assess exposures consistently across all groups: If people with higher BP (exposed cohort) are seen by their providers more frequently than those without elevated BP (nonexposed group), it also increases the chances of detecting and documenting changes in health outcomes, including CVD-related events. Therefore, it may lead to the conclusion that higher BP leads to more CVD events. This may be true, but it could also be due to the fact that the subjects with higher BP were seen more often; thus, more CVD-related events were detected and documented simply because they had more encounters with the health care system. Thus, it could bias the results and lead to an erroneous conclusion.

Question 10. Repeated exposure assessment

Was the exposure for each person measured more than once during the course of the study period? Multiple measurements with the same result increase our confidence that the exposure status was correctly classified. Also, multiple measurements enable investigators to look at changes in exposure over time, for example, people who ate high dietary sodium throughout the followup period, compared to those who started out high then reduced their intake, compared to those who ate low sodium throughout. Once again, this may not be applicable in all cases. In many older studies, exposure was measured only at baseline. However, multiple exposure measurements do result in a stronger study design.

Question 11. Outcome measures

Were the outcomes defined in detail? Were the tools or methods for measuring outcomes accurate and reliable–for example, have they been validated or are they objective? This issue is important because it influences confidence in the validity of study results. Also important is whether the outcomes were assessed in the same manner within groups and between groups.

An example of an outcome measure that is objective, accurate, and reliable is death–the outcome measured with more accuracy than any other. But even with a measure as objective as death, there can be differences in the accuracy and reliability of how death was assessed by the investigators. Did they base it on an autopsy report, death certificate, death registry, or report from a family member? Another example is a study of whether dietary fat intake is related to blood cholesterol level (cholesterol level being the outcome), and the cholesterol level is measured from fasting blood samples that are all sent to the same laboratory. These examples would get a "yes." An example of a "no" would be self-report by subjects that they had a heart attack, or self-report of how much they weigh (if body weight is the outcome of interest).

Similar to the example in Question 9, results may be biased if one group (e.g., people with high BP) is seen more frequently than another group (people with normal BP) because more frequent encounters with the health care system increases the chances of outcomes being detected and documented.

Question 12. Blinding of outcome assessors

Blinding means that outcome assessors did not know whether the participant was exposed or unexposed. It is also sometimes called "masking." The objective is to look for evidence in the article that the person(s) assessing the outcome(s) for the study (for example, examining medical records to determine the outcomes that occurred in the exposed and comparison groups) is masked to the exposure status of the participant. Sometimes the person measuring the exposure is the same person conducting the outcome assessment. In this case, the outcome assessor would most likely not be blinded to exposure status because they also took measurements of exposures. If so, make a note of that in the comments section.

As you assess this criterion, think about whether it is likely that the person(s) doing the outcome assessment would know (or be able to figure out) the exposure status of the study participants. If the answer is no, then blinding is adequate. An example of adequate blinding of the outcome assessors is to create a separate committee, whose members were not involved in the care of the patient and had no information about the study participants' exposure status. The committee would then be provided with copies of participants' medical records, which had been stripped of any potential exposure information or personally identifiable information. The committee would then review the records for prespecified outcomes according to the study protocol. If blinding was not possible, which is sometimes the case, mark "NA" and explain the potential for bias.

Question 13. Followup rate

Higher overall followup rates are always better than lower followup rates, even though higher rates are expected in shorter studies, whereas lower overall followup rates are often seen in studies of longer duration. Usually, an acceptable overall followup rate is considered 80 percent or more of participants whose exposures were measured at baseline. However, this is just a general guideline. For example, a 6-month cohort study examining the relationship between dietary sodium intake and BP level may have over 90 percent followup, but a 20-year cohort study examining effects of sodium intake on stroke may have only a 65 percent followup rate.

Question 14. Statistical analyses

Were key potential confounding variables measured and adjusted for, such as by statistical adjustment for baseline differences? Logistic regression or other regression methods are often used to account for the influence of variables not of interest.

This is a key issue in cohort studies, because statistical analyses need to control for potential confounders, in contrast to an RCT, where the randomization process controls for potential confounders. All key factors that may be associated both with the exposure of interest and the outcome–that are not of interest to the research question–should be controlled for in the analyses.

For example, in a study of the relationship between cardiorespiratory fitness and CVD events (heart attacks and strokes), the study should control for age, BP, blood cholesterol, and body weight, because all of these factors are associated both with low fitness and with CVD events. Well-done cohort studies control for multiple potential confounders.

Some general guidance for determining the overall quality rating of observational cohort and cross-sectional studies

The questions on the form are designed to help you focus on the key concepts for evaluating the internal validity of a study. They are not intended to create a list that you simply tally up to arrive at a summary judgment of quality.

Internal validity for cohort studies is the extent to which the results reported in the study can truly be attributed to the exposure being evaluated and not to flaws in the design or conduct of the study–in other words, the ability of the study to draw associative conclusions about the effects of the exposures being studied on outcomes. Any such flaws can increase the risk of bias.

Critical appraisal involves considering the risk of potential for selection bias, information bias, measurement bias, or confounding (the mixture of exposures that one cannot tease out from each other). Examples of confounding include co-interventions, differences at baseline in patient characteristics, and other issues throughout the questions above. High risk of bias translates to a rating of poor quality. Low risk of bias translates to a rating of good quality. (Thus, the greater the risk of bias, the lower the quality rating of the study.)

In addition, the more attention in the study design to issues that can help determine whether there is a causal relationship between the exposure and outcome, the higher quality the study. These include exposures occurring prior to outcomes, evaluation of a dose-response gradient, accuracy of measurement of both exposure and outcome, sufficient timeframe to see an effect, and appropriate control for confounding–all concepts reflected in the tool.

Generally, when you evaluate a study, you will not see a "fatal flaw," but you will find some risk of bias. By focusing on the concepts underlying the questions in the quality assessment tool, you should ask yourself about the potential for bias in the study you are critically appraising. For any box where you check "no" you should ask, "What is the potential risk of bias resulting from this flaw in study design or execution?" That is, does this factor cause you to doubt the results that are reported in the study or doubt the ability of the study to accurately assess an association between exposure and outcome?

The best approach is to think about the questions in the tool and how each one tells you something about the potential for bias in a study. The more you familiarize yourself with the key concepts, the more comfortable you will be with critical appraisal. Examples of studies rated good, fair, and poor are useful, but each study must be assessed on its own based on the details that are reported and consideration of the concepts for minimizing bias.

Last Updated March 2014

COGNITIVE REAPPRAISAL AND WELL-BEING

Appendix B

Results of Quality Assessment of the Included Studies Completed by both
Reviewers Using NIH Quality Assessment Tool

Author	Criteria 1	Criteria 2	Criteria 3	Criteria 4	Criteria 5	Criteria 6	Criteria 7	Criteria 8	Criteria 9	Criteria 10	Criteria 11	Criteria 12	Criteria 13	Criteria 14	**Quality**
Bienvenu et al. (2001)	Y	Y	CD	Y	N	Y	N	Y	Y	N	N	NA	CD	Y	Fair
Bienvenu et al. (2004)	N	N	CD	N	N	N	Y	Y	Y	N	Y	NA	NA	NR	Poor
Hayward et al. (2013)	Y	Y	NR	N	N	N	Y	Y	Y	N	Y	CD	NR	Y	Fair
2nd Reviewer	Y	Y	NR	N	N	N	Y	Y	Y	N	Y	N->CD	NR	Y	
Jourdy, Petot, & Aguerre (2018)	Y	Y	CD	N	N	N	Y	Y	Y	N	Y	NA	N	N	Poor
Jourdy & Petot (2017)	Y	N	CD	N	N	N	Y	Y	Y	N	Y	CD	NA	N	Poor
Naragon-Gainey & Watson (2014)	Y	N	N	Y	N	Y	Y	Y	Y	Y	N	NA	CD	N	Fair
Quilty et al. (2012)	Y	N	Y	N	N	N	N->CD	Y	Y	N	Y	NA	NA	N	Poor
2nd Reviewer	Y	N	Y	N	N	N	CD	Y	Y	N	Y	NA	NA	N	
Xia et al. (2014)	Y	Y	NR	N	N	N	N	Y	Y	N	Y	NA	NA	Y	Fair
2nd Reviewer	Y	Y	NR	N	N	N	N	Y	Y	N	Y	NA	NA	Y	
Yang et al. (1999)	Y	N	SD	Y	N	N	Y	Y	Y	N	Y	NR	NR	NR	Poor

COGNITIVE REAPPRAISAL AND WELL-BEING

Appendix C

Manuscript Submission Guidelines for *Psychological Bulletin*

Manuscript Preparation
Prepare manuscripts according to the *Publication Manual of the American Psychological Association* (6ᵗʰedition). Manuscripts may be copyedited for bias-free language (see Chapter 3 of the *Publication Manual*).

Review APA's Journal Manuscript Preparation Guidelines before submitting your article.

Double-space all copy. Other formatting instructions, as well as instructions on preparing tables, figures, references, metrics, and abstracts, appear in the *Manual*. Additional guidance on APA Style is available on the APA Style website.

Below are additional instructions regarding the preparation of display equations, computer code, and tables.

Display Equations
We strongly encourage you to use MathType (third-party software) or Equation Editor 3.0 (built into pre-2007 versions of Word) to construct your equations, rather than the equation support that is built into Word 2007 and Word 2010. Equations composed with the built-in Word 2007/Word 2010 equation support are converted to low-resolution graphics when they enter the production process and must be rekeyed by the typesetter, which may introduce errors.

To construct your equations with MathType or Equation Editor 3.0:
- Go to the Text section of the Insert tab and select Object.
- Select MathType or Equation Editor 3.0 in the drop-down menu.

If you have an equation that has already been produced using Microsoft Word 2007 or 2010 and you have access to the full version of MathType 6.5 or later, you can convert this equation to MathType by clicking on MathType Insert Equation. Copy the equation from Microsoft Word and paste it into the MathType box. Verify that your equation is correct, click File, and then click Update. Your equation has now been inserted into your Word file as a MathType Equation.

Use Equation Editor 3.0 or MathType only for equations or for formulas that cannot be produced as Word text using the Times or Symbol font.

Computer Code
Because altering computer code in any way (e.g., indents, line spacing, line breaks, page breaks) during the typesetting process could alter its meaning, we treat computer code differently from the rest of your article in our production process. To that end, we request separate files for computer code.
- In Online Supplemental Material

We request that runnable source code be included as supplemental material to the article. For more information, visit Supplementing Your Article With Online Material.

- In the Text of the Article

If you would like to include code in the text of your published manuscript, please submit a separate file with your code exactly as you want it to appear, using Courier New font with a type size of 8 points. We will make an image of each segment of code in your article that exceeds 40 characters in length. (Shorter snippets of code that appear in text will be typeset in Courier New and run in with the rest of the text.) If an appendix contains a mix of code and explanatory text, please submit a file that contains the entire appendix, with the code keyed in 8-point Courier New.

Tables
Use Word's Insert Table function when you create tables. Using spaces or tabs in your table will create problems when the table is typeset and may result in errors.

Academic Writing and English Language Editing Services
Authors who feel that their manuscript may benefit from additional academic writing or language editing support prior to submission are encouraged to seek out such services at their host institutions, engage with colleagues and subject matter experts, and/or consider several vendors that offer discounts to APA authors.

Please note that APA does not endorse or take responsibility for the service providers listed. It is strictly a referral service.

Use of such service is not mandatory for publication in an APA journal. Use of one or more of these services does not guarantee selection for peer review, manuscript acceptance, or preference for publication in any APA journal.

Submitting Supplemental Materials
APA can place supplemental materials online, available via the published article in the PsycARTICLES®database. Please see Supplementing Your Article With Online Material for more details.

Abstract and Keywords
All manuscripts must include an abstract containing a maximum of 250 words typed on a separate page. After the abstract, please supply up to five keywords or brief phrases.

Public Significance Statements
Authors submitting manuscripts to *Psychological Bulletin* are required to provide two to three brief sentences regarding the relevance or public health significance of the study or meta-analysis described in their manuscript. This description should be included within the manuscript on the abstract/keywords page.

It should be written in language that is easily understood by both professionals and members of the lay public.

Examples:
- "This meta-analysis strongly suggests that (description of a given psychosocial treatment) is an effective treatment for anxiety, but only if it is of mild to moderate severity. For persons with severe anxiety, additional treatments may be necessary."
- "This systematic review indicates that personality changes following psychotherapy and pharmacotherapy. The changes are small and persist for (description of time in months or years)"
- "This meta-analysis reveals a small to moderate effect of incidentally presenting words (e.g., as part of a game) on the actual actions of the recipients following priming. These effects are stronger when recipients of the primes are likely to value the behavior."

To be maximally useful, these statements of public significance should not simply be sentences lifted directly from the manuscript.
This statement supports efforts to increase dissemination and usage of research findings by larger and more diverse audiences. In addition, they should be able to be translated into media-appropriate statements for use in press releases and on social media.

Authors may refer to the Guidance for Translational Abstracts and Public Significance Statements page for help writing their statement.

Prior to final acceptance and publication, all public significance statements will be carefully reviewed to make sure they meet these standards. Authors will be expected to revise statements as necessary.

References
List references in alphabetical order. Each listed reference should be cited in text, and each text citation should be listed in the References section.
Examples of basic reference formats:

- **Journal Article:**
  Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin, 139,* 133–151. http://dx.doi.org/10.1037/a0028566

- **Authored Book:**
  Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.*Cambridge, MA: MIT Press.

- **Chapter in an Edited Book:**
  Gill, M. J., & Sypher, B. D. (2009). Workplace incivility and organizational trust. In P. Lutgen-Sandvik & B. D. Sypher (Eds.), *Destructive organizational*

communication: Processes, consequences, and constructive ways of organizing (pp. 53–73). New York, NY: Taylor & Francis.

Figures
Graphics files are welcome if supplied as Tiff or EPS files. Multipanel figures (i.e., figures with parts labeled a, b, c, d, etc.) should be assembled into one file. The minimum line weight for line art is 0.5 point for optimal printing.
For more information about acceptable resolutions, fonts, sizing, and other figure issues, please see the general guidelines.

When possible, please place symbol legends below the figure instead of to the side.

APA offers authors the option to publish their figures online in color without the costs associated with print publication of color figures.

The same caption will appear on both the online (color) and print (black and white) versions. To ensure that the figure can be understood in both formats, authors should add alternative wording (e.g., "the red (dark gray) bars represent") as needed.

For authors who prefer their figures to be published in color both in print and online, original color figures can be printed in color at the editor's and publisher's discretion provided the author agrees to pay:
- $900 for one figure
- An additional $600 for the second figure
- An additional $450 for each subsequent figure

Permissions
Authors of accepted papers must obtain and provide to the editor on final acceptance all necessary permissions to reproduce in print and electronic form any copyrighted work, including test materials (or portions thereof), photographs, and other graphic images (including those used as stimuli in experiments).

On advice of counsel, APA may decline to publish any image whose copyright status is unknown.
- Download Permissions Alert Form (PDF, 13KB)

Publication Policies
APA policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications.
See also APA Journals® Internet Posting Guidelines.
APA requires authors to reveal any possible conflict of interest in the conduct and reporting of research (e.g., financial interests in a test or procedure, funding by pharmaceutical companies for drug research).

- Download Disclosure of Interests Form (PDF, 38KB)
In light of changing patterns of scientific knowledge dissemination, APA requires authors to provide information on prior dissemination of the data and narrative

interpretations of the data/research appearing in the manuscript (e.g., if some or all were presented at a conference or meeting, posted on a listserv, shared on a website, including academic social networks like ResearchGate, etc.). This information (2–4 sentences) must be provided as part of the Author Note. *Psychological Bulletin* does not accept submissions which have appeared online as preprints, at this time.

Authors of accepted manuscripts are required to transfer the copyright to APA.

- For manuscripts **not** funded by the Wellcome Trust or the Research Councils UK
  Publication Rights (Copyright Transfer) Form (PDF, 83KB)

- For manuscripts funded by the Wellcome Trust or the Research Councils UK
  Wellcome Trust or Research Councils UK Publication Rights Form (PDF, 34KB)

Ethical Principles
It is a violation of APA Ethical Principles to publish "as original data, data that have been previously published" (Standard 8.13).

In addition, APA Ethical Principles specify that "after research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release" (Standard 8.14).

APA expects authors to adhere to these standards. Specifically, APA expects authors to have their data available throughout the editorial review process and for at least 5 years after the date of publication.

Authors are required to state in writing that they have complied with APA ethical standards in the treatment of their sample, human or animal, or to describe the details of treatment.

- Download Certification of Compliance With APA Ethical Principles Form (PDF, 26KB)

The APA Ethics Office provides the full Ethical Principles of Psychologists and Code of Conduct electronically on its website in HTML, PDF, and Word format. You may also request a copy by emailing or calling the APA Ethics Office (202-336-5930). You may also read "Ethical Principles," December 1992, *American Psychologist*, Vol. 47, pp. 1597–1611.

UNIVERSITY OF
EXETER

SCHOOL OF PSYCHOLOGY

DOCTORATE IN CLINICAL PSYCHOLOGY

EMPIRICAL PAPER

Using Cognitive Reappraisal and Helping Behaviour to Improve Well-being: A Single-Case Design Study

Trainee Name: **Tomas Jelinek**

Primary Research Supervisor: **Dr Pia Pechtel**

Wellcome Trust Research Fellow, University of Exeter

Secondary Research Supervisor: **Dr Nick Moberly**

Senior Lecturer, University of Exeter

Target Journal: Emotion

Word Count: 8244 words (excluding abstract, table of contents, list of figures, references, footnotes, appendices)

Submitted in partial fulfilment of requirements for the Doctorate Degree in Clinical Psychology, University of Exeter

Abstract

The use of cognitive reappraisal (CR) has been linked with improved emotional wellbeing in populations with a history of traumatic stress. Whilst research suggests that the extent to which individuals master CR (CR ability) moderates the relationship between depression and stress, studies have not attempted to improve CR ability in individuals experiencing stress due to everyday events or test for potential health benefits of this intervention. Past experimental studies using CR have largely employed group designs in which the effects of intervention are averaged across participants, leading to potentially valuable information being disguised. To this end, this study employed a single case experimental design to investigate the impact of repeated use of CR on affect, perceived stress, and depression in a female adult sample with high stress. The study also included an aspect of helping behaviour in the intervention to investigate whether there are additive benefits to using CR for self and to help others compared to using CR for oneself only.

Twelve adult females were recruited from the community (university staff and students) to take part in the study lasting 21 days. Daily measures were collected over the course of the study and pre-post study measures were taken at baseline, CR intervention, and follow-up stages. At the beginning of the intervention phase, participants were randomly allocated to one of two groups. Group 1 completed 10 days of the daily CR Task whereby daily stressors were described in writing, reappraised and then described again giving the event a newly acquired meaning. Group 2 completed five days of the CR Task followed

by five days of using CR to help reappraise written accounts of daily stressful events written by others (CR Helping Task).

At the group level, using randomisation tests, no significant improvements in emotional affect and daily stress were found in response to the intervention. At the individual level, using the reliable change index, depressive symptoms decreased reliably in three out of five participants for whom a decrease was possible and for whom CR Ability increased with the intervention. Perceived stress decreased reliably in five out of 10 participants for whom CR Ability increased. Finally, changes in depressive symptoms did not differ between groups, but, contrary to expectations, perceived stress decreased reliably in a larger number of participants in group 1 compared to group 2.

The CR intervention showed promise as a feasible short-term stand-alone intervention and demonstrated the utility of targeting specific aspects within psychological care to clarify mechanisms of change and theory. Further research is needed to explore how to optimise the intervention, particularly in terms of length and the design of the CR Helping Task.

Keywords: Reappraisal, emotion regulation, stress, depression, helping

Introduction

Emotional reactions to daily events unfold over time as a consequence of the appraisal we make of them (Folkman & Lazarus, 1985). Therefore, it is not an event that elicits a particular emotion but the person's subjective appraisal of the event that evokes an emotion. The emotional reaction triggers emotion response tendencies such as physiological responses (e.g., increased pulse) and behaviour (e.g., flight or fight; Lazarus, 1991). Gross's (1998) model of emotion regulation suggests that the most effective emotion regulation strategies are antecedent-focused strategies that are enacted even before emotion response tendencies are activated, thus avoiding the use of energy resources to manage emotional responses that have already been generated (Gross, 2001). One such antecedent-focused strategy is cognitive reappraisal (CR).

Cognitive Reappraisal and Emotional Affect

CR is an emotional regulation technique in which reinterpretation of misattributions about the emotion-eliciting event are stimulated to develop a new symbolic meaning of the experience which alters its emotional impact (Lange, Van De Ven, & Schrieken, 2003; Lazarus & Alfert, 1964). By employing CR, individuals can actively up-regulate positive affect (e.g., Krompinger, Moser, & Simons, 2008; Shiota, 2006) and down-regulate negative affect (e.g. Gross, 1998). This strategy is considered one of the key features of many psychological interventions, such as cognitive-behavioural therapy (Samoilow &

Goldfried, 2000), dialectical behavioural therapy (Lynch, Trost, Salsman, & Linehan, 2007) and psychodynamic therapy (Bateman & Fonagy, 2006). Moreover, CR has been successfully used as part of cognitive-behavioural internet-based interventions (IBI) for trauma-related mental health problems (Lange et al., 2003; Wagner, Schulz, & Knaevelsrud, 2012). Furthermore, Amstadter, Broman-Fulks, Zinzow, Ruggiero and Cercone (2009) noted that a major limitation of existing IBI packages is that they have a singular focus on specific diagnoses (e.g., depression only), which limits applicability to stress-exposed populations who are at risk of developing a broad range of mental health problems. The present study aims to rectify this limitation by recruiting a non-clinical community sample with high levels of life stress.

Cognitive Reappraisal, Perceived Stress and Depression

Due to the ability to change the course of emotional and subsequent physiological response to negative stimuli, CR has been a popular subject in the mental health coping literature for the past two decades. Research has shown that higher use of CR is observed in healthy controls compared to individuals with clinical levels of anxiety and depression (Garnefski et al., 2002; Garnefski, Teerds, Kraaij, Legerstee, & van den Kommer, 2004). Individuals who report frequent use of CR experience improved functioning in interpersonal and well-being domains, as assessed by self-reports and peer feedback (Gross & John, 2003). Frequent use of CR was also associated with better physical and mental health outcomes in patients with long-term health conditions (Moskowitz, Hult, Bussolari, & Acree, 2009) and caregivers (Pakenham, 2005). Nonetheless,

benefits of actively teaching this skill to a community sample of adults with experiences of stressful life events are yet to be demonstrated.

Cognitive Reappraisal and High Life Stress

Troy, Wilhelm, Shallcross, and Mauss (2010) found that the ability to use CR well (CR ability) moderated the relationship between intensity of life stress and depressive symptoms in a community sample of 90 women who had experienced a stressful life event in the past three months. In this study, emotional states were induced using video clips and participants' reactions were measured using physiological recordings (skin conductance), and self-report questionnaires. Results suggested that women who experienced high life stress and show low CR ability most benefit from using CR in reducing depressive symptoms. Although these results highlight important links between CR and well-being, the research did not attempt to teach or increase CR ability in women to assess improvement in well-being which may be particularly beneficial to vulnerable populations who experienced stressful life events.

CR has been successfully trialled as an IBI for Dutch adults suffering from mild to severe traumatic stress (Lange et al., 2003). Sixty-nine adults from the general population with high levels of traumatic stress completed a 5-week internet-based treatment that instructed participants to complete essays encompassing a description of the traumatic event and use of CR to challenge automatic thoughts and derive a new meaning from the events. Individuals in the treatment group reported lower trauma-related symptoms and general

psychopathology such as depression, anxiety, somatisation and sleeping problems when compared to waiting list controls. This study demonstrated that CR interventions can be successfully provided to people with a history of stressful life events using online tools. However, as the study reported a high drop-out rate, the authors suggested that future research should include an additional face-to-face element to ensure engagement and reduce attrition (Lange et al., 2003).

To the best of the author's knowledge, no previous study has sought to teach CR as a stand-alone intervention for females with a history of high life stress that tried to increase CR ability in order to alter emotional responses to everyday stressful events. Stressful life events have been consistently linked with an increase in depressive symptoms (Mazure, 1998) and an earlier onset of major depression (Hammen, 2005), suggesting populations with experiences of high life stress may be particularly vulnerable to developing mental health problems if further stressors occur. Improving the resilience of this population could have a high clinical significance. This study recruited only female participants due to the known gender differences in emotional reactivity (Charbonneau, Mezulis, & Hyde, 2009; Timmers, Fischer, & Manstead, 1998), exposure to stress (Breslau, 2002; Turner, Jay, & William, 1989), and risk for depression (Kender, Thornton, & Gardner, 2000; Nolen-Hoeksema, 2001). It also reduced heterogeneity within the sample that could limit the ability to address the study hypotheses.

To date, studies investigating the effects of CR have largely employed group designs in which the effects of intervention are averaged across

participants, which leads to potentially valuable information being obscured, for example, being able to determine who could benefit the most from a treatment. Single case experimental designs (SCED) allow individual changes in patterns in the data to be observed and can efficiently demonstrate clear causal links between intervention and behaviour change (Kazdin & Nock, 2003). Considering the findings and limitations of previous studies, the present study (a) uses CR as a stand-alone intervention in a community sample of females with high levels of life stress and low CR ability, (b) employs a SCED, (c) attempts to enhance participants' CR ability, and (d) combines online as well as in-lab face-to-face activities to limit attrition.

Finally, to increase possible benefits of the intervention, the present study adds an experimental element of helping behaviour to the CR intervention for half of the study participants. Research indicates that people who engage in helping behaviour to support others benefit from better physical and mental health outcomes (Casiday, Kinsman, Fisher, & Bambra, 2008).

Helping Behaviour and Emotional Well-being

Helping behaviour can take many forms and has been described using various terms interchangeably in the literature, including pro-social behaviour, volunteerism, and peer support (Post, 2005). This study uses the term 'helping behaviour', defined as the support that people with lived experience are able to give to one another (Mental Health Foundation, 2017). A succinct summary of the positive impact helping behaviour can have on health of the helpers as well

as those who receive it has been provided by Casiday and colleagues (2008). Their review concluded that people who engage in helping behaviour show decreased depressive symptoms, and perceived stress, and increased life satisfaction, ability to cope with own illness, and social support and interaction (Casiday et al., 2008). Therefore, adding a simple component of helping behaviour to the CR intervention could enhance its effectiveness and provide additional benefits for people in terms of well-being.

Helping Behaviour and High Life Stress

Various aspects of helping behaviour have been included as coping strategies or in clinical treatment to improve well-being (e.g., Bisson, Brayne, Ochberg, & Everly, 2007; Fallot & Harris, 2002, Jones, Roberts, & Greenberg, 2003). Midlarsky (1991) proposed five mechanisms through which engaging in helping behaviour could benefit the helper. These are (a) distraction from the person's own problems, (b) enhanced meaningfulness and purpose in life, (c) increased perception of competence and self-efficacy, (d) improved social integration, and (e) a more active lifestyle. Despite the promising evidence of the positive impact of helping behaviour, to the best of the author's knowledge, no study to date has investigated whether combining the well-established intervention of CR with helping behaviour in individuals who experienced stressful life events can have an additive effect compared to the individual CR intervention alone.

To this end, the current study examines the additive benefit of incorporating helping behaviour in a stand-alone CR intervention for females with a history of stressful life events to examine the impact on their level of depression, perceived stress and emotional well-being. Innovatively, the study combines the two intervention strategies synergistically so that the helping component can both make learning and practice of CR more effective and offer its own benefits through the mechanisms described by Midlarsky (1991).

Aims

The present study (a) investigates the impact of repeated use of CR on emotional affect, perceived stress, and depression in a female adult community sample with high life stress; and (b) investigates the additive impact of using CR to help an imaginary participant compared to using it for oneself only.

Hypotheses

1a. After CR intervention, participants will report increased positive affect, decreased negative affect and decreased daily perceived stress compared to before CR intervention.

1b. Over the course of CR intervention, participants' depressive symptoms and perceived overall stress will decrease.

2a. After CR intervention, participants who used CR intervention for themselves and to help others will show a greater increase in positive affect and greater

decreases in negative affect and daily perceived stress than people who used CR intervention for themselves only.

2b. Over the course of CR intervention, participants who used CR intervention for themselves and to help others will show greater decreases in overall perceived stress and depressive symptoms than people who used CR intervention for themselves only.

Method

SCED studies use a limited number of participants and focus on the unique differences between individuals rather than differences between groups (Rassafiani & Sahaf, 2009). By focusing on the individual, who serves as their own control, SCEDs can understand mechanisms via response patterns within treatment, allowing optimisation for individual recipients (Morgan & Morgan, 2008). Therefore, SCEDs have strong external validity making it a popular method for investigating interventions used in clinical practice. This study adhered to What Works Clearinghouse standards (WWC; Kratochwill et al., 2010) for conducting SCEDs and followed the Single-Case Reporting guideline in Behavioural Interventions (SCRIBE; Tate et al., 2016).

Design

To answer the hypotheses, a non-concurrent, randomised, multiple-baseline ABC SCED was applied. Participants completed repeated

measurements during a baseline phase (phase A), an intervention phase

(comprised of intervention 1 or 2; phase B) and a follow-up phase (phase C).

Phase A acted as a control and was compared with phases B and C.

Participants were randomly assigned to one of the two interventions. One half of

the study participants received intervention 1 (CR for self only), and the other

half received intervention 2 (CR for self and to help others). See Figure 1 for an

overview and Appendix A for a detailed description of the study design.

Phase	PHASE A Baseline					PHASE B Intervention										PHASE C Follow up					
Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Location	Lab		Flexible			Lab		Flexible				Lab		Flexible				Flexible			Lab
Measure	PRE& POST DAILY		DAILY		PRE& POST DAILY		DAILY			PRE& POST DAILY			DAILY		PRE& POST DAILY			DAILY			PRE& POST DAILY
Intervention	None					Group 1: Intervention 1 N = 6 CR Task for 10 days										None					
						Group 2: Intervention 2 N = 6 CR Task for 5 days + CR Helping Task for 5 days															

Figure 1. Study diagram with a baseline length of five days

Participants

Twelve female participants with a recent history of life stress were

recruited from a non-clinical population using posters and an online research

article (Appendix B). Ages ranged from 19 to 57 (Mean = 38.4, SD = 11.4).

Interested participants were emailed an information sheet and invited to

complete a screening survey (Appendix C). Those who met inclusion criteria

were contacted via phone, the study design was explained, and university

appointments arranged. Eight participants were university staff, two were

postgraduate, and two were undergraduate students. Participants were given £40 as a gift for participation.

Inclusion criteria. Candidates were required to (a) have experienced at least two stressful life events in the past 12 months and considered these to have had a negative impact on their lives, (b) be female, (c) 18 years of age or older, (d) able to access the internet from home, (e) able to use keyboard or touchpad to type, (f) commit to completing daily internet-based tasks lasting between 10 and 30 minutes for the duration of 21 days, and (g) be willing to attend four face-to-face appointments with the researcher.

Exclusion criteria. To increase the probability of participants' CR ability being low, participants were excluded if they had received cognitive behavioural therapy (CBT) in the past six months and/or had changed medication for a mental health condition in the past six weeks. CBT includes training and practice of emotion regulation strategies such as CR, which would limit the effect of the intervention provided by this study. Recent changes in medication can cause changes in mood and this could further confound the study results.

Measures and Materials

Life Experiences Survey (LES; Sarason, Johnson, & Siegel, 1978; Appendix D). Cumulative stress was measured using section 1 of the LES, which includes 47 questions assessing a wide range of potentially stressful events that an individual may have experienced in the past 12 months. Respondents indicate whether they experienced each event and the kind of

impact the event had on them. The LES also allows respondents to add up to three unique stressful events that may not have been mentioned among the 47 items. The respondent is asked to rate the impact of each event on a 7-point scale ranging from "*extremely negative*" to "*extremely positive*". Only the negative impact of stressful life events was used as a measure of cumulative stress as negative events have been found to better predict negative psychological outcomes (Sarason, Sarason, Potter, & Antoni, 1985). A high score on the LES (range 0 – 150) indicates high cumulative stress resulting from negatively appraised adverse events.

Daily outcome measures (see Appendix E).

The Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988). The PANAS is a 20-item self-report measure of positive and negative state affect. The scale asks respondents to indicate on a 1 to 5 Likert scale how they are feeling at the present moment in relation to 20 different markers of positive or negative affect. Positive affect (PA) is represented by the extent to which a subject experiences pleasurable engagement with the environment (e.g., enthusiasm), while high subjective distress and unpleasurable engagement is indicative of negative affect (NA; e.g., lethargy). The reliabilities (internal consistencies) of the PANAS positive affect and negative affect scales were described using Cronbach's alpha and estimated at .89 for the PA scale, and .85 for the NA scale (Crawford & Henry, 2004).

Daily Stress measure. As a brief measure of daily stress, participants were asked to respond to the question: "Over the past 24 hours, how stressed did you feel overall?" using a 10-point scale (where 1 was '*not at all stressed*' and 10 was '*extremely stressed*'). This measure was designed specifically for this study.

Frequency of CR Use measure. As a daily measure to assess the frequency of CR use, participants were asked to respond to the question "Over the past 24 hours, how often did you use cognitive reappraisal?" using a 10-point scale (where 1 was '*not at all*' and 10 was '*all the time*'). This measure was designed specifically for this study.

Pre and post-study outcome measures (Appendix F). All pre and post-study outcome measures were administered at five measurement times (MT): beginning and end of baseline (A), middle and end of treatment (B), and end of follow-up (C) (see Figure 1).

The Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001). The PHQ-9 is a nine-item questionnaire assessing the frequency of symptoms of depression over the previous two weeks. Items are scored 0 ("*not at all*") to 3 ("*nearly every day*") with a maximum total score of 27. A total score of 15 or higher may be indicative of moderate depression. The PHQ-9 has good test-retest reliability (over a four-week period; correlation coefficient of .84) and internal consistency (Cronbach's alpha of .89) (Kroenke, Spitzer, & Williams, 2001).

The Perceived Stress Scale (PSS-10; Cohen, Kamarck, &

Merhelstein, 1994)*.* The PSS-10 is used to measure perception of the degree

to which situations in one's life are stressful. The scale, comprised of 10 items,

was designed for use in community samples and focuses on stressful feelings

and thoughts during the last month. The PSS-10 has good reliability coefficients

(Cronbach's alpha = .91; Mitchell, Crane, & Kim, 2008). Test-retest reliability

yielded correlations ranging from .55 (six-week interval) to .61 (12-month

interval) (Cohen, Kamarck, & Mermelstein, 1983; Cole, 1999).

As the length between intervals of the administration of PHQ-9 and PSS

was less than what the two measures stipulate, the instructions were re-

phrased to reflect the time since last completing the measures.

Emotion Regulation Questionnaire (ERQ; Gross & John, 2003). The

ERQ is a self-report 7-point Likert scale questionnaire measuring habitual use

of two emotional regulation strategies: cognitive reappraisal and expressive

suppression. Only the cognitive reappraisal subscale was used in this study as

a measure of CR Ability. It has 6 items and measures the extent to which

respondents attach a positive meaning to stressful events in terms of personal

growth. The subscale has been shown to have good validity and reliability in a

large community sample as demonstrated by Cronbach's alpha of .89 (Preece,

Becerra, Robinson, & Gross, 2019).

Intervention

The CR intervention was specifically developed for the purposes of this study and comprised four key elements: (a) education, (b) practice with support, (c) repeated independent practice, and (d) feedback and skill consolidation. To control for dose response effects, both intervention 1 and intervention 2 lasted exactly 10 days for each participant.

Intervention 1: Cognitive reappraisal (Appendix G). At the beginning of treatment (phase B), all participants were introduced to CR by the researcher in a face-to-face appointment (appointment 2). CR was explained as a technique to help change the meaning of a stressful event to improve individuals' emotional reaction to the event (Troy et al., 2010).

To learn and practice CR, participants were presented with a written scenario that included an example use of CR. The researcher discussed the example with each participant to ensure they understood CR. Participants were then asked to write about a stressful event that happened to them in the past 24-48 hours, apply CR and re-write the event using a computer. The researcher then discussed the completed CR scenario to ensure the participant understood the concept and was able to use CR sufficiently to complete the subsequent CR tasks on their own.

Participants independently completed the CR task daily (at flexible times) over a 10-day period. After five days, participants attended a face-to-face appointment and were given the opportunity to discuss issues that they had encountered. The researcher also provided verbal feedback on the participant's

use of CR to date. Finally, participants were asked to complete another CR task in the appointment, and support was offered.

Intervention 2: Cognitive reappraisal with helping behaviour (Appendix H). Intervention 2 comprised five days of intervention 1 followed by five days of using a variation of CR with a focus on helping others (CR helping task). The CR helping task asked participants to use CR to reappraise a description of a stressful event that the participant was led to believe was written by another participant in the study. Therefore, instead of using CR to rewrite own descriptions of own stressful events, participants were asked to apply CR to rewrite others' experiences in order to help them learn the skill and experience the emotional benefit of using CR for others. To ensure participants' descriptions of stressful events remained confidential, the scenarios were written by the researcher in partnership with a member of the Lived Experience Group (LEG) within the Mood Disorders Centre at the University of Exeter. The scenarios were approved to be realistic representations of stressful situations. Similar to intervention 1, participants were asked to complete the first CR helping task in a face-to-face appointment (after five days of intervention 1) and support was offered where appropriate to complete the exercise.

Procedure

The study used Qualtrics computer software (Qualtrics, 2017) to collect all data. The study was conducted in a naturalistic setting with participants practising the CR tasks and completing outcome measures at a time and place

suitable to them and fitting in with their daily routines. Blinding was not possible as the researcher knew the unique participant numbers that had been ascribed to individuals.

The study involved four face-to-face appointments up to 90 minutes long with the researcher at the Exeter University (see Figure 1). At appointment 1 on the first day of phase A, the study was outlined, the risk protocol explained, and informed consent given (Appendix I). Participants were explained the basic concept of CR and asked to complete daily and pre and post-study measures on the computer. The Mood Disorders Centre Protocol for Assessing and Reporting Risk (Appendix J) was followed if participants answered ≥1 on the suicidality and self-harm question of the PHQ-9.

After the first face-to-face appointment (day 1), participants completed daily measures (PANAS, daily stress, frequency of CR use) for another 20 consecutive days. The researcher contacted participants by email if daily measures were missed, and by phone if participant scored ≤ 9 on the daily stress scale for two consecutive days to ask about well-being, monitor risk and signpost to services where appropriate. Pre and post-study measures (PSS, PHQ-9) were completed at appointment 1 and online at the end of phase A.

The second face-to-face appointment was scheduled for the first day of phase B (day 5, 6 or 7) after completing phase A. The CR intervention began at this appointment and the CR task was practised independently by participants for five days. Pre and post-study measures were completed online for the third time on the fifth day of phase B.

The third face-to-face appointment took place on the sixth day of phase B. Participants were asked to reflect on their experience of practising the CR task and issues raised were addressed. Knowledge of the correct use of CR was monitored and support was given if needed. Participants were randomly assigned to receive intervention 1 or intervention 2. Scripts used for instructions and training elements were identical for the two interventions except for variations in the focus of CR (self vs. others). Pre and post-study measures were completed online for the fourth time on the tenth day of phase B.

Intervention 1 and 2 ended after 10 days marking the end of phase B. Participants were told they could keep using CR if they had found it helpful. Participants then completed phase C (5, 6 or 7 days) and met the researcher for a final, fourth appointment at the end of phase C (day 21). A fifth set of pre and post-study measures was completed, and debriefing information and £40 gift for participation were provided (Appendix K). Participants were also offered to receive a summary of the study results when they became available.

Piloting

The study design, intervention and materials were developed in consultation with a member of the LEG. A pilot study with one participant was then trialled prior to the full empirical study. Feedback from LEG and the pilot participant was incorporated into the main study. Feedback largely related to the survey flow and accessibility of language. Ethical approval was gained from the University of Exeter Ethics Committee (Appendix L).

Data Analysis

Data analysis was completed using SPSS 25 (IBM Corporation, 2017), R open source statistical environment (R Core Team, 2019) and Microsoft Excel. A repeated measures ANOVA was used to check whether the intervention significantly changed participants' Frequency of CR Use and CR Ability.

Construction of visual displays using traditional visual analysis method is an essential part of SCED data analysis (Morley, 2017). Following guidelines on conducting SCEDs (Kratochwill et al., 2010; Tate et al., 2016), this study supplemented traditional visual analysis with statistical analyses as these approaches are considered complementary (Maggin & Odom, 2014) and their use in combination is becoming increasingly popular in the SCED literature (Maggin, O'Keeffe, & Johnson, 2011). To this end, to answer hypothesis 1a, randomisation tests (Onghena & Edgington, 2005) were used to assess significant differences in daily outcome measures between phases using all 12 participants. Visual analysis was also used to explore changes between phases within individual participants. To answer hypothesis 2a, differences between groups receiving different interventions was assessed using visual analysis only. Hypotheses 1b and 2b were answered by assessing reliable change (Jacobson & Truax, 1991) between phases and comparing the two groups with reference to reliable change. Appendix M provides a detailed description and justification of all methods of data analysis used in this study.

Results

All 12 participants completed the 21-day study. No procedural changes were required. Compliance was excellent, with only seven daily recording opportunities missed across all participants (participant 1 and 4 missed one day each, participant 5 missed five days). Missing data points are not plotted on raw data plots. Mean cumulative stress scores measured by the LES was 13.5 (*SD* = 7.9). Three participants (2, 4, and 9) scored less than the average score (Mean = 8.3, *SD* = 6.3) of a non-clinical normative sample reported by Denisoff and Endler (2000).

Table 1 shows mean scores of each phase for CR Ability and Frequency of CR Use. Mean CR Ability score at baseline was low compared to an average score of a normative sample (Mean = 6.6, *SD* = 1.0) reported by Gross and John (2003). Both Frequency of CR Use and CR Ability significantly improved between baseline and follow-up. Please refer to Appendix N for a detailed description of results of repeated measures ANOVA used in this part of analysis.

Table 1

Mean Score per each Phase for Frequency of CR Use and CR Ability

	Frequency of CR Use			CR Ability		
Partici pant	Phase A	Phase B	Phase C	Phase A	Phase B	Phase C
1	1.8	1.8	1.7	5	3.0	4.3
2	4.5	6.3	3.7	2.5	5.3	5.5
3	2.4	6.7	9.5	3	5.8	6.2
4	2.3	6	5.7	4.3	5	5
5	1	5.2	6.5	3.3	3	4.3
6	1	2.9	4	4.5	4.3	4.8
7	4.3	3.1	2.4	4.3	2.7	3.3
8	5	5.6	6.6	4.3	5	5.3
9	2	5	3	4.3	6	6.7
10	1	2.6	1.5	2	2.2	2.7
11	1	1.9	1	4.3	5	5.5
12	2.8	3.6	3.8	4.3	5	5
Group	2.4	4.2	4.1	3.8	4.4	4.9

Positive Affect

The intervention aimed to increase positive affect (PA) scores. Figure 2 shows plotted raw PA data for all 12 participants. Participants 1 to 6 received intervention 1 and participants 7 to 12 received intervention 2.

COGNITIVE REAPPRAISAL AND WELL-BEING

Figure 2. Raw positive affect data for individual participants

Within phase evaluation.

Stability and level. Evaluation of the baseline phase using the stability criterion indicated data were stable in three participants (4, 5, and 6) in group 1 and five participants (7, 8, 9, 10, and 12) in group 2. Compared to an average score of a normative sample (Mean = 30.6, *SD* = 7.9) reported by Crawford and Henry (2004), evaluation of level within baseline using the median indicated baseline PA was extremely low in two participants (10 and 11) in group 2, relatively low in five participants (1, 3, 4, 5, and 6) in group 1 and two participants (7 and 8) in group 2, and about average in one participant (2) in group 1 and three participants (8, 9, and 12) in group 2. Further detailed analysis of within phase PA data is described in Appendix O.

Between phases evaluation.

Changes in trend and level. Among participants whose baseline data were stable, evaluation of changes in trend between phases indicated PA went from a deteriorating trend in baseline to a stable or improving trend during intervention in participants 7 and 8 (group 2). Level of PA however changed between phases in a counter-therapeutic manner for both participants. Comparing intervention and follow-up, PA continued to improve in an accelerating trend for both participants 7 and 8, however, analysis of level change did not show improvement. No other therapeutic changes in trend were observed between phases for the other three participants (4, 5, and 6) in group 1 and three participants (9, 10, and 12) in group 2 whose baseline data was stable.

Overlap. Overlap between two phases was evaluated using the non-overlap of all pairs index (NAP; Parker & Vannest, 2009) calculation for each participant as well as all 12 participants combined. Overall, the intervention did not have a desired effect on PA (NAP = .36). On an individual level, the intervention had a medium effect on two participants (2 and 3) and a weak effect on one participant (6) in group 1 and a weak effect on one participant (10) in group 2.

Randomisation test. Monte-Carlo simulation indicated the CR intervention did not have a significant effect on PA (p = .997).

Summary. At a group level, PA did not significantly improve for the 12 participants following intervention. Based on the results of visual analysis, no differences between groups 1 and 2 were observed. On an individual level, changes in trend in a therapeutic direction between baseline and intervention were observed in participants 7 and 8, however level of data decreased between phases and NAP calculation indicated data largely overlapped. Finally, although participants 2 and 3 demonstrated a medium effect, their baseline data was not stable, consequently causal inferences made from any observable patterns in the data are unreliable.

Negative Affect

The intervention aimed to decrease negative affect (NA) scores following baseline. Figure 3 shows plotted raw NA data for all 12 participants.

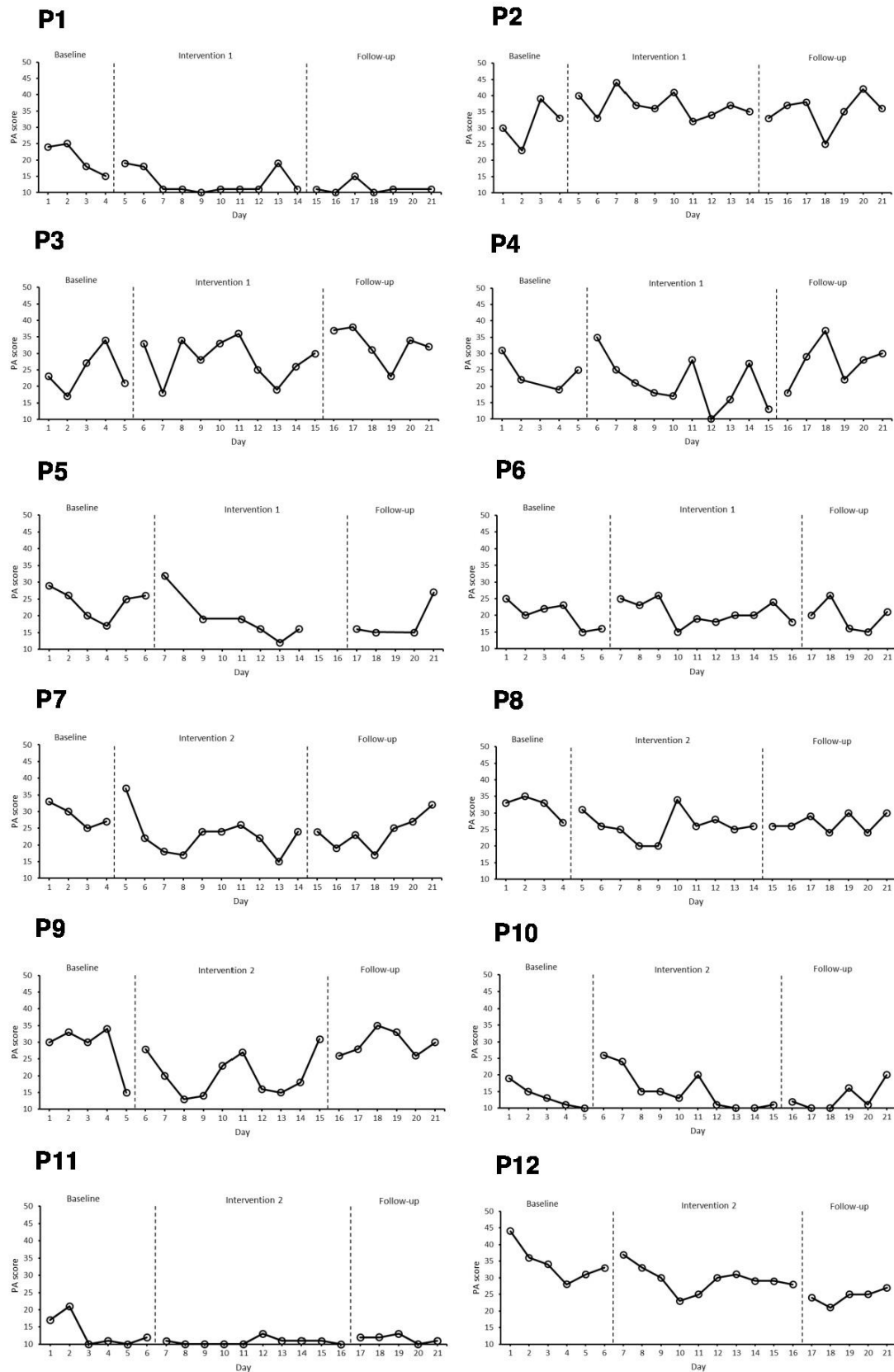COGNITIVE REAPPRAISAL AND WELL-BEING
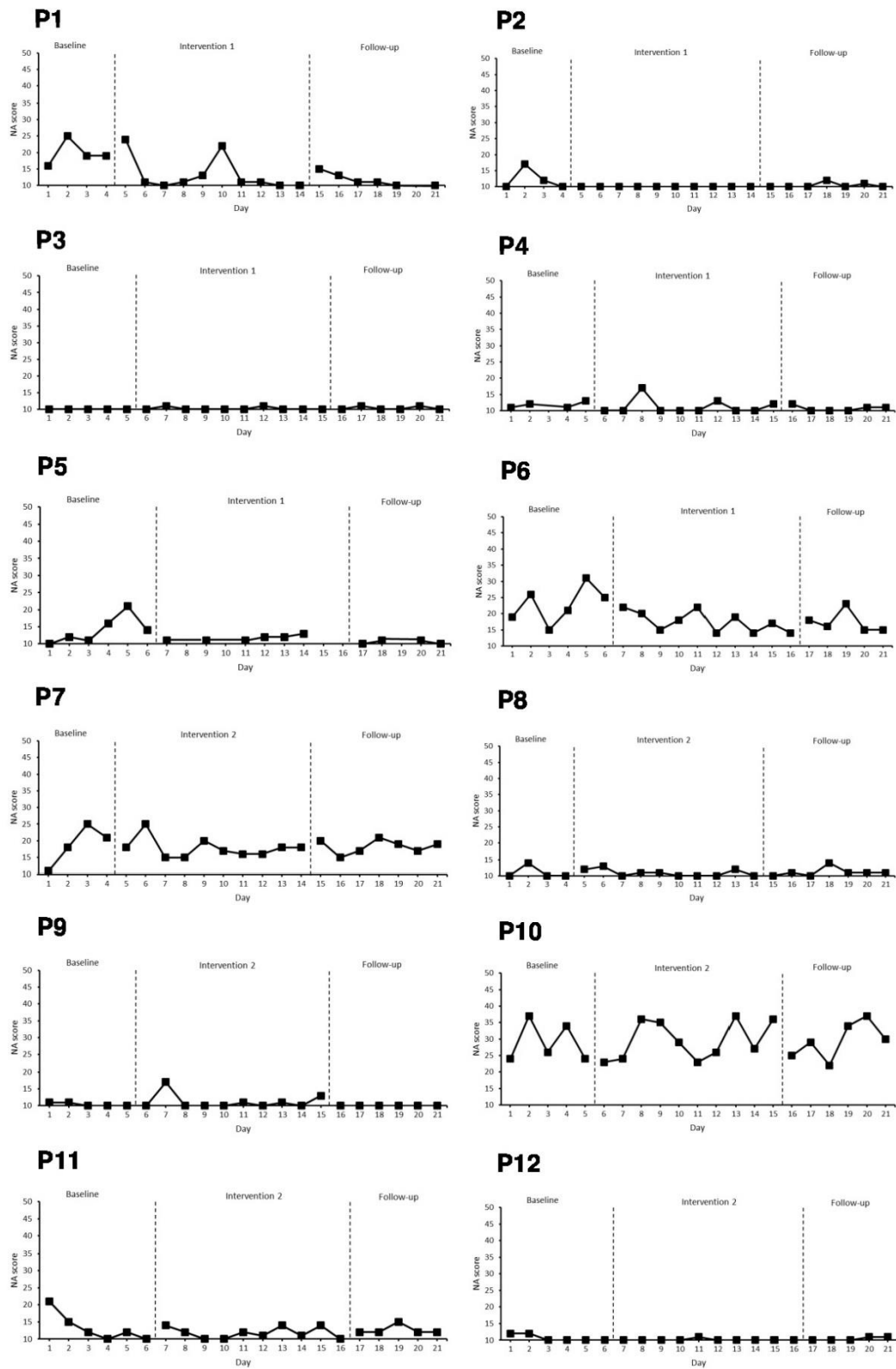
Figure 3. Raw negative affect data for individual participants

Within phase evaluation.

Stability and level. Evaluation of the baseline phase indicated data were stable in three participants (3, 4, and 5) in group 1 and three participants (9, 11, and 12) in group 2. Compared to an average score of a normative sample (Mean = 16.7, *SD* = 6.4) reported by Crawford and Henry (2004), evaluation of level within baseline using the median indicated baseline NA was extremely low in three participants (2, 3, and 4) in group 1 and four participants (8, 9, 11, and 12) in group 2, relatively low in one participant (5) in group 1, about average in one participant (1) in group 1 and one participant (7) in group 2 and relatively high in one participant (6) in group 1 and one participant (10) in group 2. Further detailed analysis of within phase NA data is presented in Appendix O.

Between phases evaluation.

Changes in trend and level. Among participants whose baseline data were stable, evaluation of changes in trend between phases indicated NA went from a deteriorating trend in baseline to a stable or improving trend during intervention only in participant 4 (group 1). Level of NA also slightly decreased in a therapeutic direction between the two phases in this participant. In participant 5 (group 1), trend did not change from deteriorating between phases, however level shifted slightly in a therapeutic direction. Therapeutic shifts in level and/or trend were difficult to detect in participants 3 (group 1), 9, 11, and 12 (group 2) as their baseline level of NA was already extremely low. Baseline data in participant 6 (group 1) were not stable, however, data stabilised in intervention and both trend and level indicated therapeutic effect. Comparing intervention and follow-up, in participant 4 level of NA remained low. In

participant 5, trend stabilised, and level remained low. In participant 6, trend remained improving and level further shifted in a therapeutic direction.

Overlap. Overlap in NA data between two phases was evaluated using the NAP calculation. Overall the intervention had a weak desired effect on NA (NAP = .62). On an individual level, the intervention had a medium effect on five participants (1, 2, 4, 5, and 6) in group 1 and a weak effect on four participants (7, 9, 11, and 12) in group 2.

Randomisation test. Monte-Carlo simulation indicated the CR intervention did not have a significant effect on NA ($p = .38$).

Summary. At a group level, NA did not significantly improve for the 12 participants following the intervention. Based on the results of visual analysis, no differences between groups 1 and 2 were observed. On an individual level, analysis of level, trend and overlap indicated improvements in participants 4 and 5; however, these were extremely small. Furthermore, baseline NA was already extremely low in a large number of participants, which prevented any potential effect of the intervention to be demonstrated. Finally, NA baseline data were largely unstable in participants with higher baseline levels of NA (1, 2, 6, 7, and 10), which made any causal inferences from observable patterns in the data unreliable.

Daily Stress

The intervention aimed to decrease Daily Stress (DS) scores following baseline. Figure 4 shows plotted raw DS data for all 12 participants.

Within phase evaluation.

Stability and level. Evaluation of the baseline phase indicated data were stable in three participants (1, 5, and 6) in group 1 and three participants (8, 9, and 10) in group 2. Evaluation of level within baseline indicated DS of 6 and above was present in four participants (1, 4, 5, and 6) in group 1 and four participants (7, 9, 10, and 11) in group 2. Further detailed analysis of within phase DS data is described in Appendix O.

Between phases evaluation.

Changes in trend and level. Among participants whose baseline data were stable, evaluation of changes in trend between phases indicated DS went from a deteriorating trend in baseline to a stable or improving trend during intervention only in participants 5 (group 1) and 7 (group 2). DS level did not change for participant 5 but improved for participant 7. In participants 1 (group 1) and 9 (group 2), DS level changed in a therapeutic manner and trend changed from level to improving. Several other participants demonstrated therapeutic changes in level (3, 4, 11, and 12), however, these could have been observed due to an already decelerating, improving trend. No other therapeutic changes in trend were observed between phases for the other three participants (6, 8, and 10; group 2) whose baseline data was stable. Improvements in DS were largely maintained in follow-up for participants 1, 5, 7, and 9.

COGNITIVE REAPPRAISAL AND WELL-BEING
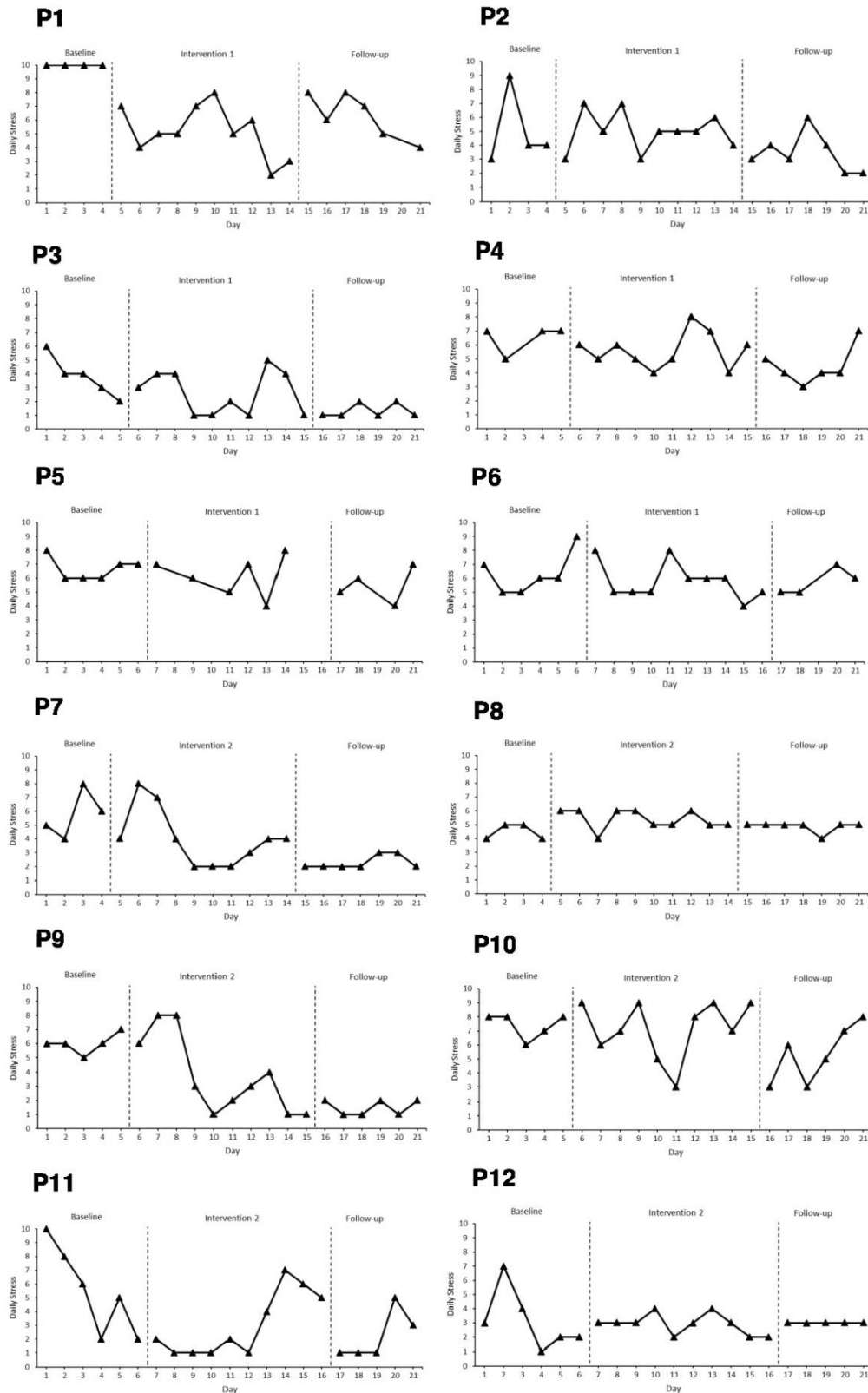
Figure 4. Raw Daily Stress data for individual participants

Overlap. Overlap in DS data between two phases was evaluated using the NAP calculation. Overall the intervention had a moderate effect on Daily Stress (NAP = .68). On an individual level, the intervention had a weak effect on two participants (2 and 6), a moderate effect on three participants (3, 4, and 5) and a strong effect on one participant (1) in group 1 and a weak effect on one participant (10) and a moderate effect on three participants (7, 9, and 11) in group 2.

Randomisation test. Monte-Carlo simulation indicated the CR intervention did not have a significant effect on DS ($p = .45$).

Summary. At a group level, DS did not significantly improve following intervention. Based on the results of visual analysis, no differences between groups 1 and 2 were observed. At an individual level, changes in trend and/or level in a therapeutic direction between baseline and intervention were observed in two (5 and 9) out of five participants for whom baseline data was stable and Frequency of CR Use and CR Ability increased. NAP calculation indicated a moderate effect of the intervention on DS in these two participants. Several other participants showed therapeutic changes in level of DS, however, these could have been attributed to an already improving trend in baseline or general variability in data that was demonstrated within phases. Finally, the two participants (1 and 7) for whom Frequency of CR Use and CR Ability did not increase showed improvements in DS, which renders causal inferences made from data patterns observed in the other two participants less reliable.

Depressive Symptoms

The intervention aimed to decrease depressive symptoms as measured by the Patient Health Questionnaire (PHQ-9). Table 2 details PHQ-9 scores for individual participants in phases A, B, and C. Mean baseline PHQ-9 score for all 12 participants was 6.3 (SD = 5.6), slightly higher than the average score of a normative sample (Mean = 3.3, SD = 3.8) reported by Kroenke et al., (2001). Seven participants were largely asymptomatic at baseline (PHQ-9 score ≤ 4), which meant any potential effect of the intervention would not have shown using this measure. Of those whose scores were of clinical significance, two participants (4 and 6) in group 1 and one participant (11) in group 2 showed reliable change (RCI = 3.9; Kroenke et al., 2001) and one participant (1) in group 1 and one participant (10) in group 2 did not. No differences between groups were noted.

COGNITIVE REAPPRAISAL AND WELL-BEING

Table 2

Reliable Change for Depressive Symptoms

Participant	Phase A PHQ-9 score	Phase B PHQ-9 score	Phase C PHQ-9 score	Reliable Change? A -> B	Reliable Change? B -> C	Reliable Change? A -> C
1	8	10	6	No	Yes	No
2	2	1	1	No	No	No
3	4	2	1	No	No	No
4	7	8	3	No	Yes	Yes
5	4	4	4	No	No	No
6	13	8	9	Yes	No	Yes
7	4	5	3	No	No	No
8	2	2	2	No	No	No
9	2	0	0	No	No	No
10	19	21	20	No	No	No
11	11	7	5	Yes	No	Yes
12	0	0	0	No	No	No

Perceived Stress

The intervention aimed to decrease perceived stress as measured by the Perceived Stress Scale (PSS). Table 3 details PSS scores for individual participants in phases A, B, and C. Mean baseline PSS score for all 12 participants was 23.8 ($SD = 8.1$) indicating high stress levels compared to an average score of a normative data (Mean = 13.7, $SD = 6.6$) reported by Cohen et al. (1994). Reliable change (RCI = 8.37; Cohen et al., 1994) occurred

COGNITIVE REAPPRAISAL AND WELL-BEING

between baseline and intervention in three participants (3, 4, and 5) in group 1 and one participant (9) in group 2. Comparing baseline and follow-up, reliable change occurred in four participants (2, 3, 4, and 5) in group 1 and one participant (9) in group 2. Overall, PSS baseline scores decreased in five out of 12 participants, either during intervention or at follow-up. The majority of reliable changes occurred in group 1 compared to group 2.

Table 3

Reliable Change for Perceived Stress

Participant	Phase A PSS score	Phase B PSS score	Phase C PSS score	Reliable Change? A -> B	Reliable Change? B -> C	Reliable Change? A -> C
1	27	27	21	No	No	No
2	14	6	4	No	No	Yes
3	26	4	4	Yes	No	Yes
4	32	23	17	Yes	No	Yes
5	26	17	17	Yes	No	Yes
6	18	17	17	No	No	No
7	18	22	15	No	No	No
8	11	12	10	No	No	No
9	16	7	3	Yes	No	Yes
10	36	36	30	No	No	No
11	23	23	20	No	No	No
12	11	6	6	No	No	No

Discussion

The purpose of this study was two-fold. First, it examined the impact of the repeated use of CR on affect, stress, and depression in a non-clinical sample of adult females with high levels of life stress. Second, the study investigated the additive impact on affect, stress and depression of using CR for oneself and to help others compared to using CR just for oneself. Specifically, it was predicted that the CR intervention would increase PA and decrease NA, and DS. It was also predicted that depressive symptoms and perceived stress would decrease over the course of treatment. Finally, it was predicted that these effects would be greater for those participants who use CR for themselves and to help others (intervention 2) compared to participants using CR only for themselves (intervention 1).

With regards to hypotheses 1a and 2a, at the group level, no significant improvements in PA, NA, or DS were found in response to the intervention. In addition, no differences between intervention 1 and 2 were observed. At the individual level, results were encouraging in three participants who demonstrated therapeutic changes in DS. However, one participant did not report an increase in Frequency of CR Use or CR Ability in response to the intervention, which suggests their improvement in DS cannot be attributed to the use of CR specifically. Consequently, hypotheses 1a and 2a were rejected.

With regards to hypotheses 1b and 2b, depressive symptoms decreased reliably in three out of five participants for whom a decrease in PHQ-9 score was possible and for whom Frequency of CR Use and CR Ability increased with the intervention. Perceived stress decreased reliably in five out of 10

participants for whom Frequency of CR Use and CR Ability increased.

Depressive symptoms did not seem to improve more in one group compared to

the other, but perceived stress decreased reliably in a larger number of

participants who used CR for self only compared to those who used CR for self

and others. These findings suggest that, for those who use CR and their ability

improves, the intervention may have a positive impact on depressive symptoms

and perceived stress. Finally, using CR for oneself only appears more effective

in reducing perceived stress than using it for self and to help others.

Despite the well-established evidence from previous studies that links the

use of CR to improved emotional wellbeing (e.g., Krompinger et al., 2008),

participants did not demonstrate a significant improvement in affect in response

to the CR intervention. Several factors may have contributed to this result.

First, baseline PA and NA data were unstable in nearly half of all study

participants, which may have impacted on the reliability of any causal

inferences based on observed data patterns. Whilst this would have impacted

on the feasibility of the study, a longer baseline allowing baseline data within

each participant to stabilise before introducing the intervention may have

improved the design and provided more confidence in the effects of the

intervention (Lane & Gast, 2014).

Second, previous literature suggested that CR is most effective for

improving negative affect (Gross & John, 2003). Through appraising the

stressor in a more helpful way, people respond to stressful situations with more

equanimity thus increased resilience and reduced NA (Ray, McRae, Ochsner, &

Gross, 2010). In the current study baseline NA was extremely low, which could

have prevented any potential therapeutic effect of the intervention being demonstrated. Although stress tends to correlate with NA (Cohen, Tyrrell, & Smith, 1993), it did not seem to be the case for this study sample, which was largely comprised of university staff and postgraduate students. It is possible that participants in this study were relatively high functioning and their mood was not particularly affected by high stress.

A similar issue with participants' low baseline scores pertains to the PHQ-9 used to measure depression. Although this meant that effects of the intervention could not be demonstrated in several participants, three out of five participants who demonstrated clinically significant levels of depression at baseline benefited from daily use of CR. These findings are in line with previous theory (Folkman & Lazarus, 1985) and empirical evidence (Lange et al., 2003) suggesting that identification and challenge of unhelpful automatic thoughts resulting from stressful experiences can result in an acquisition of a more positive meaning (Lange et al., 2003).

Participants did not demonstrate a significant reduction in daily stress over the course of the CR intervention, although data showed a therapeutic trend. One potential cause of the non-significant result is the variability of baseline data in half of the study participants. Perceived stress reliably changed for five participants over the course of the study. These benefits are particularly meaningful in light of the increased risk of developing depression in individuals experiencing high levels of stress (Hammen, 2005; Mazure, 1998). CR was reported to down-regulate negative emotions resulting from negative appraisals of stressful events and therefore break the vicious cycle of negative emotions

leading to further stress (Shiota, 2006). The current findings support the utility of using CR as a stand-alone intervention for preventative purposes in non-clinical populations high in stress.

Contrary to predictions based on previous findings (Casiday et al., 2008), adding helping behaviour to the CR intervention did not have an additive benefit. Notably, reappraising stressful situations described by others proved rather challenging for participants. Qualitative differences between reappraisals that participants completed in the third appointment and reappraisals completed in the four subsequent days suggested more support and guided practice may have been needed. Participants frequently reported not being sure "*what to tell them*" and had doubts about the quality and usefulness of their reappraisals. In addition, some participants guessed that the scenarios had been written by the researcher and this would have turned the helping task into a mere practice exercise. Indeed, despite best intentions, it seemed the laboratory-based helping task lacked social interaction and feedback and therefore was ineffective in providing its benefits through the mechanisms reported by Midlarsky (1991). These mechanisms propose that people benefit from helping largely due to improvements in social integration and sense of purpose. In fact, where participants struggled, felt responsible or incompetent, the task may have been counter-productive and this might explain the low number of participants in group 2 for whom perceived stress reliably changed compared to group 1.

Finally, Post (2005) stresses the importance of the altruistic aspect of supportive behaviours. He posits that the biggest benefits of helping behaviour are seen where help is given voluntarily and is motivated by concern for others

rather than by anticipation of reward. Indeed, intrinsically motivated behaviour has been found to elicit more positive emotions than extrinsically motivated behaviour (Ryan & Deci, 2000). The design of the helping task was not successful in incorporating this important aspect, partially because the laboratory circumstances in which it took place tend to be non-altruistic by nature (Rachlin, 2002). In short, participants were likely to take part in the study for their own benefit, rather than the benefit of others. Despite this, altruistic behaviour has been demonstrated in laboratory experiments in the past through gift-giving computer games, group activities and other means (for a review, see Andreoni, Harbaugh, & Vesterlund, 2008). The helping task could be improved by adopting some of these methods.

Clinical Implications

CR was tested as a stand-alone treatment. Testing individual elements that can be used in more complex treatment packages has its utility. For instance, when comparing outcomes between therapeutic treatments and treatment as usual, mechanisms of change may remain unclear. Demonstrating the usefulness of individual treatment elements such as CR can help clarify theory about what might be the key mechanisms of change in more complex care packages. Consequently, the results are promising as they support the utility of including CR among the active ingredients of psychological interventions aimed to reduce depressive symptoms (Samoilow & Goldfried, 2000). Yet, future work is needed to overcome some of the methodological limitations in the current design to further explore this argument.

Overall, the study was successful in improving low CR ability and demonstrated that participants will keep using CR skilfully even after the intervention ends. The intervention is potentially cost-effective as it is relatively easy to administer and combines internet-based tasks with regular face-to-face meetings, which could take place in low-intensity or general health settings. This strategy also drastically reduces attrition, a frequently cited issue in internet-based interventions (Lange et al., 2003).

Although not all participants benefited from reduced stress and depressive symptoms, the study offers a promising, cost-effective tool that could be used in low-intensity or health settings to help vulnerable populations to gain CR ability, a skill that is widely recognised to benefit health (Samoilow & Goldfried, 2000).

Whilst further work is needed to refine the intervention and test for effectiveness, it could potentially serve as a preventative strategy to reduce the incidence of mental health disorders requiring expensive treatments within services - an important cause identified in clinical research (Arango et al., 2018).

Strengths and Limitations

A particular strength of the study was the use of a non-concurrent, randomised, multiple-baseline SCED that involved randomisation tests, visual analysis and measures of reliable change. This allowed a thorough analysis of group effects as well as changes occurring within individuals. Furthermore, the

incorporation of a follow-up phase enabled observations of effects of the intervention after its discontinuation thus giving some indication of the longer-term benefits. The ratio of internet-based tasks and face-to-face appointments was helpful in limiting drop-out and missing data points. Finally, the relatively short length of the intervention combining internet-based tasks and face-to-face contact makes it highly feasible and clinically relevant, both key considerations in designing studies testing novel interventions (National Institute for Health Research, 2011).

The study also had several limitations. The limited number of phase changes did not allow for the use of randomisation tests on individual participants. Furthermore, whilst the study adhered to the minimum requirements for SCED baseline length (Kratochwill et al., 2010), baseline PA, NA, and DS data did not have sufficient time to stabilise before the introduction of the intervention in a large number of participants, thus potentially confounding the results. Related to this, stress was measured by a single-item questionnaire, which in itself may have caused variability in the data due to measurement error. The study also aimed to recruit a community sample, but all participants were affiliated with the university, therefore likely to be high functioning. This could have contributed to the low levels of NA and depressive symptoms and impacts on the generalisability of the study's findings.

Future Research

Qualitative feedback on the intervention could be sought from study participants and used for improving face validity of the helping task and other optimisation. The length of the intervention could be extended or, if possible, offered in optimal length for each participant so that their CR ability would reach highest possible levels. If this was unfeasible, extending the length of follow-up would allow researchers to assess whether CR Ability continues to improve, and benefits of the intervention are sustained after the intervention has ended. To further progress the evidence base for CR interventions, future research could focus on recruiting subclinical samples with higher levels of stress and depression severity as participants with these characteristics seemed to have benefitted the most from the CR intervention used in this study. Future SCED studies may need to carefully balance feasibility and methodological improvements pertaining to recruitment, complexity of daily measures, stability of baseline data, and larger number of phase changes.

Conclusions

This study findings offer limited evidence in support of the utility of a stand-alone emotion regulation intervention for an at-risk sample of females. Using a single case experimental design, this study tested the benefits of repeated use of CR on affect, stress and depression. Possible additive benefits of using CR for self and to help others were also explored. Although the intervention increased participants' CR ability, positive impact demonstrated by

a reliable decrease in depressive symptoms and stress was limited to only some participants. Lack of stability in baseline data and low levels of NA and depressive symptoms could explain benefits not being demonstrated in other participants. Using CR for self only more beneficial compared to using it for self and to help others. The CR intervention showed promise as a feasible short-term stand-alone intervention and demonstrated the utility of targeting specific aspects within psychological care to clarify mechanisms of change and theory. Further research is needed to explore how to optimise the intervention, particularly in terms of length and the design of the helping task.

References

Amstadter, A. B., Broman-Fulks, J., Zinzow, H., Ruggiero, K. J., & Cercone, J. (2009). Internet-based interventions for traumatic stress-related mental health problems: a review and suggestion for future research. *Clinical Psychology Review, 29*, 410-420. doi:10.1016/j.cpr.2009.04.001

Andreoni, J., Harbaugh, W. T., & Vesterlund, L. (2008). Altruism in experiments. *The New Palgrave Dictionary of Economics, 1–8*, 134-138. doi:10.1057/9780230226203.0035

Arango, C., Díaz-Caneja, C. M., McGorry, P. D., Rapoport, J., Sommer, I. E., Vorstman, J. A., ... & Carpenter, W. (2018). Preventive strategies for mental health. *The Lancet Psychiatry. 5*, 591-604. doi:10.1016/s2215-0366(18)30057-9

Bateman, A., & Fonagy, P. (2006). *Mentalization-based treatment for borderline personality disorder: A practical guide*. OUP Oxford. doi:10.1093/med/9780198570905.001.0001

Bisson, J. I., Brayne, M., Ochberg, F. M., & Everly Jr, G. S. (2007). Early psychosocial intervention following traumatic events. *American Journal of Psychiatry, 164*, 1016-1019. doi:10.1176/ajp.2007.164.7.1016

Breslau, N. (2002). Gender differences in trauma and posttraumatic stress disorder. *The Journal of Gender-specific Medicine, 5*, 34-40. doi:10.1016/s1550-8579(06)80194-4

Bulté, I., & Onghena, P. (2013). The Single-Case Data Analysis Package: Analysing Single-Case Experiments with R Software, *Journal of Modern Applied Statistical Methods, 12*, 450-478. doi:10.22237/jmasm/1383280020

Casiday, R., Kinsman, E., Fisher, C., & Bambra, C. (2008). Volunteering and health: what impact does it really have. *London: Volunteering England*, *9*, 1-13. doi:10.1017/s1041610208007096

Charbonneau, A. M., Mezulis, A. H., & Hyde, J. S. (2009). Stress and emotional reactivity as explanations for gender differences in adolescents' depressive symptoms. *Journal of Youth and Adolescence*, *38*, 1050-1058. doi:10.1007/s10964-009-9398-8

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*, 385-396. doi:10.2307/2136404

Cohen, S., Kamarck, T., & Mermelstein, R. (1994). Perceived stress scale. *Measuring stress: A guide for health and social scientists*, *10*, 235-283. doi:10.1037/t02889-000

Cohen, S., Tyrrell, D. A., & Smith, A. P. (1993). Negative life events, perceived stress, negative affect, and susceptibility to the common cold. *Journal of Personality and Social Psychology*, *64*, 131-140. doi:10.1037//0022-3514.64.1.131

Cole, S. R. (1999). Assessment of differential item functioning in the Perceived

    Stress Scale-10. *Journal of Epidemiology and Community Health*, *53*,

    319–320. doi:10.1136/jech.53.5.319

Crawford, J. R., & Henry, J. D. (2004). The positive and negative affect

    schedule (PANAS): construct validity, measurement properties and

    normative data in a large non-clinical sample. *The British Journal of*

    Clinical Psychology, 43, 245-265. doi:10.1348/0144665031752934

Duff, K. (2012). Evidence-based indicators of neuropsychological change in the

    individual patient: relevant concepts and methods. *Archives of Clinical*

    Neuropsychology, *27*, 248-261. doi:10.1093/arclin/acr120

Edgington, E., & Onghena, P. (2007). *Randomization tests*. Chapman and

    Hall/CRC. doi:10.1111/j.1541-0420.2007.00905_10.x

Fallot, R. D., & Harris, M. (2002). The Trauma Recovery and Empowerment

    Model (TREM): Conceptual and practical issues in a group intervention

    for women. *Community Mental Health Journal*, *38*, 475-485.

    doi:10.1037/e648702010-001

Ferron, J., & Sentovich, C. (2002). Statistical power of randomization tests used

    with multiple-baseline designs. *The Journal of Experimental Education*,

    70, 165-178. doi:10.1080/00220970209599504

Folkman, S., & Lazarus, R. S. (1985). If it changes it must be a process: Study

    of emotion and coping during three stages of a college examination.

Journal of Personality and Social Psychology, *48*, 150-170. doi:10.1037//0022-3514.48.1.150

Garnefski, N., Van Den Kommer, T., Kraaij, V., Teerds, J., Legerstee, J., & Onstein, E. (2002). The relationship between cognitive emotion regulation strategies and emotional problems: comparison between a clinical and a non-clinical sample. *European Journal of Personality*, *16*, 403-420. doi:10.1002/per.458

Garnefski, N., Teerds, J., Kraaij, V., Legerstee, J., & van den Kommer, T. (2004). Cognitive emotion regulation strategies and depressive symptoms: Differences between males and females. *Personality and Individual Differences*, *36*, 267-276. doi:10.1016/s0191-8869(03)00083-7

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2,* 271-299. doi:10.1037//1089-2680.2.3.271

Gross, J. J. (2001). Emotion regulation in adulthood: Timing is everything. *Current Directions in Psychological Science*, *10*, 214-219. doi:10.1111/1467-8721.00152

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85,* 348-362. doi:10.1037/0022-3514.85.2.348

Hammen, C. (2005). Stress and depression. *Annual Review of Clinical Psychology, 1*, 293-319. doi:10.1146/annurev.clinpsy.1.102803.143938

Heyvaert, M., Moeyaert, M., Verkempynck, P., Van den Noortgate, W., Vervloet, M., Ugille, M., & Onghena, P. (2017). Testing the intervention effect in single-case experiments: A Monte Carlo simulation study. The Journal of Experimental Education, *85*, 175-196. doi:10.1080/00220973.2015.1123667

Heyvaert, M., & Onghena, P. (2014). Analysis of single-case data: Randomisation tests for measures of effect size. *Neuropsychological Rehabilitation*, *24*, 507-527. doi:10.1080/09602011.2013.818564

IBM Corporation (2017). IBM SPSS Statistics for Windows (Version 25.0) [Computer software]. Available from https://www.qualtrics.com/uk/?rid=ip&prevsite=en&newsite=uk&geo=GB &geomatch=uk

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19. doi:10.1037//0022-006x.59.1.12

Jamieson, M., Cullen, B., McGee-Lennon, M., Brewster, S., & Evans, J. J. (2014). The efficacy of cognitive prosthetic technology for people with memory impairments: A systematic review and meta-analysis. *Neuropsychological Rehabilitation*, *24*, 419-444. doi:10.1080/09602011.2013.825632

Jones, N., Roberts, P., & Greenberg, N. (2003). Peer group risk assessment: a post-traumatic management strategy for hierarchical organizations. *Occupational Medicine*, *53*, 469-475. doi:10.1093/occmed/kqg093

Jones, R. R., Weinrott, M. R., & Vaught, R. S. (1978). Effects of serial dependency on the agreement between visual and statistical inference. *Journal of Applied Behavior Analysis, 11*, 277-283. doi:10.1901/jaba.1978.11-277

Kazdin, A. E. (1982). Single-case experimental designs in clinical research and practice. *New Directions for Methodology of Social & Behavioral Science*, *13*, 33–47. doi:10.1016/j.brat.2018.11.015

Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry*, *44*, 1116-1129. doi:10.1111/1469-7610.00195

Kendler, K. S., Thornton, L. M., & Gardner, C. O. (2000). Stressful life events and previous episodes in the etiology of major depression in women: an evaluation of the "kindling" hypothesis. *American Journal of Psychiatry*, *157*, 1243-1251. doi:10.1176/appi.ajp.157.8.1243

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). What Works Clearinghouse: Single-case designs technical documentation. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kroenke, K., Spitzer, R. L., & Williams, W. (2001). The PHQ-9: Validity of a brief

    depression severity measure. *Journal of General Internal Medicine*, *16*,

    606-616. doi:10.1046/j.1525-1497.2001.016009606.x

Krompinger, J. W., Moser, J. S., & Simons, R. F. (2008). Modulations of the

    electrophysiological response to pleasant stimuli by cognitive

    reappraisal. *Emotion*, *8*, 132-137. doi:10.1037/1528-3542.8.1.132

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental

    design studies: Brief review and guidelines. *Neuropsychological*

    Rehabilitation, *24*, 445-463. doi:10.1080/09602011.2013.815636

Lange, A., Van De Ven, J. P., & Schrieken, B. (2003). Interapy: treatment of

    post-traumatic stress via the internet. *Cognitive Behaviour Therapy*, *32*,

    110-124. doi:10.1080/16506070302317

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American*

    Psychologist, *46*, 352-367. doi:10.1037//0003-066x.46.4.352

Lazarus, R. S., & Alfert, E. (1964). Short-circuiting of threat by experimentally

    altering cognitive appraisal. *The Journal of Abnormal and Social*

    Psychology, *69*, 195. doi:10.1037/h0044635

Lynch, T. R., Trost, W. T., Salsman, N., & Linehan, M. M. (2007). Dialectical

    behavior therapy for borderline personality disorder. *Annual Review of*

    *Clinical Psychol*ogy, *3*, 181-205.

    doi:10.1146/annurev.clinpsy.2.022305.095229

Maggin, D. M., & Odom, S. L. (2014). Evaluating single-case research data for

    systematic review: A commentary for the special issue. *Journal of School*

    Psychology, *52*, 237–241. doi:10.1016/j.jsp.2014.01.002

Maggin, D. M., O'Keeffe, B. V., & Johnson, A. H. (2011). A quantitative

    synthesis of methodology in the meta-analysis of single-subject research

    for students with disabilities: 1985–2009. *Exceptionality*, *19*, 109-135.

    doi:10.1080/09362835.2011.565725

Mazure, C. M. (1998). Life stressors as risk factors in depression. *Clinical*

    Psychology: Science and Practice, *5*, 291-313. doi:10.1111/j.1468-

    2850.1998.tb00151.x

Mental Health Foundation (2017). Peer Support [Blog post]. Retrieved from

    https://www.mentalhealth.org.uk/a-to-z/p/peer-support

Midlarsky, E. (1991). Helping as coping. Prosocial Behavior: *Review of*

    Personality and Social Psychology, *12*, 238–264.

    doi:10.1037/e504482011-001

Mitchell, A. M., Crane, P. A., & Kim, Y. (2008). Perceived stress in survivors of

    suicide: Psychometric properties of the perceived stress scale. *Research*

    in Nursing and Health, 31, 576-585. doi:10.1002/nur.20284

Morgan, D. L., & Morgan, R. K. (2008). *Single-case research methods for the*

    behavioral and health sciences. Sage publications.

    doi:10.4135/9781483329697

Morley, S. (2017). *Single case methods in clinical psychology.* Oxon, UK:

Routledge.

Moskowitz, J. T., Hult, J. R., Bussolari, C., & Acree, M. (2009). What works in

coping with HIV? A meta-analysis with implications for coping with

serious illness. *Psychological Bulletin*, *135*, 121-141.

doi:10.1037/a0014210

National Institute for Health Research (2011, May). Points to consider when

assessing the feasibility of research. Retrieved from

http://www.netscc.ac.uk/FAQs/PDFs/AMRC-MCRN-Feasibility-Guide-

May2011.pdf

Nolen-Hoeksema, S. (2001). Gender differences in depression. *Current

Directions in Psychological Science, 10*, 173-176. doi:10.1111/1467-

8721.00142

Onghena, P., & Edgington, E. S., (2005). Customization of pain treatments:

single-case design and analysis. *Clinical Journal of Pain, 21*, 56-68.

doi:10.1097/00002508-200501000-00007

Pakenham, K. I. (2005). Relations between coping and positive and negative

outcomes in carers of persons with multiple sclerosis (MS). *Journal of

Clinical Psychology in Medical Settings*, *12*, 25-38. doi:10.1007/s10880-

005-0910-3

Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single case

research: Nonoverlap of all pairs (NAP). *Behavior Therapy, 40*, 357-367.

doi:10.1016/j.beth.2008.10.006

Parsonson, B. S. & Baer, D. M. (1986). The graphic analysis of data. In A.

Poling & R. W. Fuqua (Eds.), *Research methods in applied behaviour

analysis: Issues and advances* (pp. 157–186). doi:10.1007/978-1-4684-

8786-2_8

Post, S. G. (2005). Altruism, happiness, and health: It's good to be good.

International Journal of Behavioral Medicine, 12, 66-77.

doi:10.1207/s15327558ijbm1202_4

Preece, D. A., Becerra, R., Robinson, K., & Gross, J. J. (2019). The Emotion

Regulation Questionnaire: Psychometric properties in general community

samples. *Journal of Personality Assessment*, *21*, 1-9.

doi:10.1080/00223891.2018.1564319

Qualtrics (2017). [Web-based application]. Provo, Utah, USA. Retrieved from

http://www.qualtrics.com

R Core Team (2019). R: A language and environment for statistical computing.

R Foundation for Statistical Computing, Vienna, Austria. [Computer

software]. Available from http://www.R-project.org/.

Rachlin, H. (2002). Altruism and selfishness. *Behavioral and Brain Sciences*,

25, 239-250. doi:10.1017/s0140525x02000055

Rassafiani, M., & Sahaf, R. (2010). Single case experimental design: an

overview. *International Journal of Therapy and Rehabilitation*, *17*, 285-

289. doi:10.12968/ijtr.2010.17.6.48151

Ray, R. D., McRae, K., Ochsner, K. N., & Gross, J. J. (2010). Cognitive

reappraisal of negative affect: Converging evidence from EMG and self-

report. *Emotion, 10*, 587-592. doi:10.1037/a0019015

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic

definitions and new directions. *Contemporary Educational Psychology,*

25, 54-67. doi: 10.1006/ceps.1999.1020

Samoilow, A., & Goldfried, M. R. (2000). Role of emotion in cognitive-behavior

therapy. *Clinical Psychology: Science and Practice*, *7*, 373–385.

doi:10.1093/clipsy.7.4.373

Sarason, I. G., Johnson, J. H., Siegel, J. M. (1978). Assessing the impact of life

changes: Development of the Life Experiences Survey. *Journal of*

Consulting and Clinical Psychology, *46*, 932-946.

Sarason, I. G., Sarason, B. R., Potter, E. H., & Antoni, M.H. (1985). Life events,

social support, and illness. *Psychosomatic Medicine, 47*, 156–163.

doi:10.1037/e517512009-001

Shiota, M. N. (2006). Silver linings and candles in the dark: Differences among

positive coping strategies in predicting subjective well-being. *Emotion*, *6*,

335. doi:10.1037/1528-3542.6.2.335

Tate, R. L., Perdices, M., Rosenkoetter, U., McDonald, S., Togher, L., Shadish, W., ... & Sampson, M. (2016). The Single-Case Reporting Guideline In Behavioural Interventions (SCRIBE) 2016: Explanation and elaboration. *Archives of Scientific Psychology, 4*, 10–31. doi:10.1037/arc0000027

Tate, R. L., Perdices, M., Rosenkoetter, U., Wakim, D., Godbee, K., Togher, L., & McDonald, S. (2013). Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: The 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. *Neuropsychological Rehabilitation, 23*, 619-638. doi:10.1080/09602011.2013.824383

Timmers, M., Fischer, A. H., & Manstead, A. S. (1998). Gender differences in motives for regulating emotions. *Personality and Social Psychology Bulletin, 24*, 974-985. doi:10.1177/0146167298249005

Troy, A. S., Wilhelm, F. H., Shallcross, A. J., & Mauss, I. B. (2010). Seeing the silver lining: cognitive reappraisal ability moderates the relationship between stress and depressive symptoms. *Emotion, 10*, 783-795. doi:10.1037/a0020262

Turner, R. J., & Avison, W. R. (1989). Gender and depression: assessing exposure and vulnerability to life events in a chronically strained population. *Journal of Nervous and Mental Disease. 177*, 443-455. doi:10.1097/00005053-198908000-00001

Urbaniak, G. C., & Plous, S. (2013). Research Randomizer (Version 4.0) [Computer software]. Available from http://www.randomizer.org/

Vannest, K.J., Parker, R.I., Gonen, O., & Adiguzel, T. (2016). Single Case

   Research: web based calculators for SCR analysis. (Version 2.0) [Web-

   based application]. College Station, TX: Texas A&M; University.

   Retrieved from

   https://web.archive.org/web/20180404183208/http://www.singlecaserese

   arch.org/calculators/nap

Wagner, B., Schulz, W., & Knaevelsrud, C. (2012). Efficacy of an Internet-based

   intervention for posttraumatic stress disorder in Iraq: a pilot study.

   Psychiatry Research, 195, 85-88. doi:10.1016/j.psychres.2011.07.026

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of

   brief measures of positive and negative affect: the PANAS scales.

   Journal of Personality and Social Psychology, *54,* 1063-1070.

   doi:10.1037//0022-3514.54.6.1063

Appendix A

Detailed Description of the Study Design

Simulation analysis carried out by Ferron and Sentovich (2002) showed that sufficient power (>.80) is achieved in SCEDs with at least 20 repeated measurements and four participants. According to the WWC standards, a phase must have a minimum of three data points to qualify as an attempt to demonstrate an effect (Kratochwill et al., 2010). Furthermore, incorporating randomisation of participants into different lengths of baseline improves internal validity and statistical conclusion validity of single-case designs (Edgington & Onghena, 2007). This study was expected to meet the outlined power criteria as it (a) collected > 20 repeated measurements, (b) for 12 participants, (c) each phase included > 3 data points, and (d) participants were randomly assigned to different baseline lengths and to the variations in treatment (interventions 1 or 2). In total, the SCED protocol included 21 measurement times (MT) comprised of a minimum of four and maximum of six MTs for phase A, ten MTs for phase B and a minimum of five and maximum of seven MTs for phase C. The length of the study was 21 days for each participant (e.g., 5+10+6 as shown in Figure A1, p. 127).

The non-concurrent and staggered beginning of treatment is required to ensure that the randomisation is sufficiently powered because the greater the number of possible phase changes, the less likely that any improvement will be randomly associated with a particular phase change. The design also enables the assessment and ruling out of historical confounding variables (Heyvaert &

Onghena, 2014). All subjects were randomised a priori, using an online research randomisation software (Research Randomiser; Urbaniak & Plous, 2013). The decision to limit the length of baseline (between 4 and 6 days) and therefore the number of possible phase changes was made in order to limit participant fatigue and increase the feasibility of the study as lack of face-to-face contact and the burden of describing stressful events were among the factors that contributed to high drop-out rates in previous CR research (Lange et al., 2003). See Figure A1 for the randomised moment of phase change for each participant.

To investigate the hypotheses, the study utilised an ABC design to test the effect of a CR intervention (phase B) on affect, daily stress, depressive

Figure A1. Sequence of A, B, and C phase for individual participants

Participant	Intervention	Sequence
1	1	AAAA*BBBBBBBBBB*CCCCCCC
2	1	AAAA*BBBBBBBBBB*CCCCCCC
3	1	AAAAA*BBBBBBBBBB*CCCCCC
4	1	AAAAA*BBBBBBBBBB*CCCCCC
5	1	AAAAAA*BBBBBBBBBB*CCCCC
6	1	AAAAAA*BBBBBBBBBB*CCCCC
7	2	AAAA*BBBBBBBBBB*CCCCCCC
8	2	AAAA*BBBBBBBBBB*CCCCCCC
9	2	AAAAA*BBBBBBBBBB*CCCCCC
10	2	AAAAA*BBBBBBBBBB*CCCCCC
11	2	AAAAAA*BBBBBBBBBB*CCCCC
12	2	AAAAAA*BBBBBBBBBB*CCCCC

symptoms and perceived stress after baseline (phase A). The multiple-baseline AB sequence enables causal inferences to be made between intervention and behaviour. Using the staggered moment of phase change for each subject, sustained improvements in outcome measures following the introduction of the intervention would suggest that these improvements are due to the intervention rather than maturation, spontaneous remission, selection and other threats to internal validity (Kratochwill et al., 2010; Tate et al., 2013). Each participant in the study served as their own control. Comparisons of phases B and C allows the study to establish whether the effect of the intervention is sustained when the intervention is discontinued, an important aspect of research testing novel clinical interventions (Morley, 2017).

As the study also examined if using CR to help others provides an added benefit, two interventions each lasting a total of 10 days were studied following baseline (intervention 1 = CR Task; intervention 2 = CR Task plus CR Helping Task). In that part of the study, phase B and C are combined and compared with phase A. The two interventions are compared using visual analysis of data in groups 1 and 2. Participants were randomly assigned one of the interventions so that half of the study participants received intervention 1 and the other half received intervention 2.

Appendix B

Recruitment Poster

UNIVERSITY OF EXETER

STRESS STUDY

Do you have recent experience with stressful life events?

Can you access the internet from home?

Are you interested in finding new ways of coping with stress?

Researchers at the University of Exeter are running a study to understand whether practicing a skill that allows us to change the meaning we make of stressful events can improve the way we feel. We are also curious whether using this skill for others can help us even more.

The following eligibility criteria apply to the study:

✓ **Are female aged 18 or over**

✓ **Experienced stressful life events in the past 18 months**

✓ **Can access internet from home via desktop computer, laptop, or tablet**

✓ **Able to complete questionnaires and/or short computer-based exercises lasting 10 to 30 minutes each day for the duration of 21 days (flexible, from home)**

✓ **Willing to attend 4 appointments at Exeter University lasting up to 90 minutes**

✗ **Must not have changed medication for a mental health issue in the past six weeks**

✗ **Must not have received cognitive behavioural therapy in the past six months**

What does taking part involve?

Participation involves completing online questionnaires and/or short exercises using your desktop, laptop or tablet every day for 21 consecutive days. You will also be asked to attend four appointments at the university which will last up to 90 minutes. In the exercises we will ask you to think of an every-day (minor) stressful event that happened to you in the past 48 hours, describe it in writing and practice changing the meaning you make of it. We will teach you how to do this at the beginning of the study. **You will be reimbursed £40 for your time and travel**. If you would like more information about the study or are interested in taking part, please contact:

Tomas Jelinek
E-mail: tj285@exeter.ac.uk

The study has received approval from the Ethics Committee
at the University of Exeter (reference: eCLESPsy000606)

COGNITIVE REAPPRAISAL AND WELL-BEING

Appendix C

Online Screening Survey

Q1. Hi, welcome to screening. Before you continue to the screening questionnaire, please enter below the unique study code you have been emailed by the researcher together with the link to this survey.

> 260612

Q2. Great, thank you. Please proceed to the screening questionnaire.

Q3. Dear participant,

Thank you for your interest in the study and welcome to screening. The study aims to find out whether a repeated daily use of cognitive reappraisal (assigning a more positive meaning to a stressful event) of daily stressful events can reduce stress and improve your emotional well-being. We are also interested to see whether these benefits are even bigger if you use cognitive reappraisal to help another person.

Before we can offer you a place in the study, however, we need to make sure the intervention we offer has a good chance to work for you. We also need to make sure we do not cause any harm.

In order to do that we ask you to complete some questionnaires. There are no right or wrong answers. Your answers will be kept completely confidential.

The screening survey that we ask you to complete today can have three different outcomes:

1. You will meet the inclusion criteria and we will offer you a place in the study.

2. You will not meet the inclusion criteria in which case we will not be able to offer you a place, but we will offer you to be added to the Exeter University psychology participant pool. This will allow you to take part in any other studies that you might be notified of in the future.

3. You will not meet the inclusion criteria and your answers indicate you may be experiencing mental health difficulties. If this happens we might advise you to contact your GP and/or offer to do this on your behalf. If we need to contact a health professionals for additional support, we will always inform you first. We will also offer to speak with you on the phone and discuss the above options.

Please note it might take up to five working days to receive the result of the screening.

Would you like to proceed to the screening questionnaires?

Remember, you can withdraw at any point of the screening.

- ◉ YES, proceed
- ○ NO, I changed my mind and do not wish to take part in the study

Q4. Are you sure? If you select YES, we will not be able to offer you a place in the study.

COGNITIVE REAPPRAISAL AND WELL-BEING

Q1. Hi, welcome to screening. Before you continue to the screening questionnaire, please enter below the unique study code you have been emailed by the researcher together with the link to this survey.

260612

Q2. Great, thank you. Please proceed to the screening questionnaire.

Q3. Dear participant,

Thank you for your interest in the study and welcome to screening. The study aims to find out whether a repeated daily use of cognitive reappraisal (assigning a more positive meaning to a stressful event) of daily stressful events can reduce stress and improve your emotional well-being. We are also interested to see whether these benefits are even bigger if you use cognitive reappraisal to help another person.

Before we can offer you a place in the study, however, we need to make sure the intervention we offer has a good chance to work for you. We also need to make sure we do not cause any harm.

In order to do that we ask you to complete some questionnaires. There are no right or wrong answers. Your answers will be kept completely confidential.

The screening survey that we ask you to complete today can have three different outcomes:

1. You will meet the inclusion criteria and we will offer you a place in the study.

2. You will not meet the inclusion criteria in which case we will not be able to offer you a place, but we will offer you to be added to the Exeter University psychology participant pool. This will allow you to take part in any other studies that you might be notified of in the future.

3. You will not meet the inclusion criteria and your answers indicate you may be experiencing mental health difficulties. If this happens we might advise you to contact your GP and/or offer to do this on your behalf. If we need to contact a health professionals for additional support, we will always inform you first. We will also offer to speak with you on the phone and discuss the above options.

Please note it might take up to five working days to receive the result of the screening.

Would you like to proceed to the screening questionnaires?

Remember, you can withdraw at any point of the screening.

- ⦿ YES, proceed
- ◯ NO, I changed my mind and do not wish to take part in the study

Q4. Are you sure? If you select YES, we will not be able to offer you a place in the study.

COGNITIVE REAPPRAISAL AND WELL-BEING

Q11. During the study, are you willing to attend 4 appointments at the Department of Psychology at Exeter University lasting up to 90 minutes?

(Exact times and dates will be arranged with the researcher.)

◉ YES
○ NO

Q12.
A stressful life event is an event with a distinct starting point that had a significant, negative impact on your life.

Some examples of a stressful life event could be sudden unemployment, illness, injury, death of a close family member or friend, long distance move, exposure to crime, the end of a long-term romantic relationship and many others...

How many stressful life events have you experienced over the past **12** months?

◉ None
○ One
○ Two
○ Three
○ More than three

Q13. On the gauge below where 0 is no negative impact and 10 is extremely negative impact, please indicate the extent to which you view the stressful life events you have experienced in the past 18 months as having negative impact on your life.

This question was not displayed to the respondent.

Q14. Thank you!

You have completed the screening questionnaires. Please allow up to five working days to receive the result of the screening.

When we get in touch, you will receive one of the three different results:

1. You will meet the inclusion criteria and we will offer you a place in the study. If this happens we will send you more information about the study, speak with you on the phone and arrange a first appointment with you at the university should you choose to enrol in the study.

2. You will not meet the inclusion criteria in which case we will not be able to offer you a place, but we will offer you to be added to the Exeter University psychology participant pool. This will give you an opportunity to participate in future studies you might be invited for.

3. You will not meet the inclusion criteria and your answers indicate you may be experiencing mental health difficulties. If this happens we might advise you to contact your GP and/or offer to do this on your behalf. If we need to contact a health professionals for additional support, we will always inform you first. We will also offer to speak with you on the phone and discuss the above options.

This question was not displayed to the respondent.

Q15.

Please Keep Safe

If you experienced any emotional difficulties during the completion of the screening questionnaires and wish to speak with a professional, please consider contacting one of the following agencies:

1. Depression and Anxiety Service
Call: 01392 675630
E-mail: dpn-tr.ExeterDAS@nhs.net
More information: https://www.dpt.nhs.uk/our-services/depression-and-anxiety-das

2. The Samaritans
Free to call number: 01392 116 123
E-mail: jo@samaritans.org
More information: https://www.samaritans.org/branches/samaritans-exeter-mid-east-devon

Finally, if you experienced any technical difficulties with the screening surveys, please contact the researcher via email on: tj285@exeter.ac.uk

This question was not displayed to the respondent.

Q16. Thank you. This is the end of the screening survey. Please click NEXT to finish.

This question was not displayed to the respondent.

Appendix D

Life Experiences Survey (Sarason, Johnson, & Siegel, 1979)

Q10.

The Life Experiences Survey

Listed below are a number of events which sometimes bring about change in the lives of those who experience them and which necessitate social readjustment. For each of the events, please:

1. indicate the time period during which you have experienced each event (indicate not applicable if you have never experienced this event or the event happened more than 1 year ago)

2. indicate the extent to which you viewed the event as having either a positive or negative impact on your life at the time the event occurred.

	Not applicable	0 to 6 months	7 to 12 months	Extremely negative	Moderately negative	Somewhat negative	No impact	Slightly positive	Moderately positive	Extremely positive
Marriage	✔									
Detention in jail or comparable institution	✔									
Death of spouse	✔									
Major change in sleeping habits (much more or much less sleep)		✔		✔						
Death of close family member: a. mother	✔									
b. father	✔									
c. brother	✔									
d. sister	✔									
e. grandmother	✔									
f. grandfather	✔									
g. other	✔									

Q11. Continued (2/6)

	Not applicable	0 to 6 months	7 to 12 months	Extremely negative	Moderately negative	Somewhat negative	No impact	Slightly positive	Moderately positive	Extremely positive
Major change in eating habits (much more or much less food intake)		✔		✔						
Foreclosure on mortgage or loan	✔									
Death of close friend		✔		✔						
Outstanding personal achievement		✔								✔
Minor law violations (traffic tickets, disturbing the peace, etc.)	✔									
Male: Wife/girlfriend's pregnancy	✔									
Female: Pregnancy	✔									
Changed work situation (different work responsibility, major change in working conditions, working hours, etc.)		✔		✔						✔
New job	✔									

Q12. Continued (3/6)

	Not applicable	0 to 6 months	7 to 12 months	Extremely negative	Moderately negative	Somewhat negative	No impact	Slightly positive	Moderately positive	Extremely positive
Serious illness or injury of close family member:	✔									
a. father	✔									
b. mother	✔									
c. sister	✔									
d. brother	✔									
e. grandfather	✔									
f. grandmother	✔									
g. spouse	✔									
h. other	✔									
Sexual difficulties	✔									

	Not applicable	0 to 6 months	7 to 12 months	Extremely negative	Moderately negative	Somewhat negative	No impact	Slightly positive	Moderately positive	Extremely positive
Trouble with employer (in danger of losing job, being suspended, demoted, etc.)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐

Q13.
Continued (4/6)

	Not applicable	0 to 6 months	7 to 12 months	Extremely negative	Moderately negative	Somewhat negative	No impact	Slightly positive	Moderately positive	Extremely positive
Trouble with inlaws	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Major change in financial status (a lot better off or a lot worse off)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Major change in closeness of family members (increased or decreased closeness)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Gaining a new family member (through birth, adoption, family member moving in, etc.)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Change of residence	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Marital separation from mate (due to conflict)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Major change in church activities (increased or decreased attendance)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Marital reconcilation with mate	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Major change in number of arguments with spouse (a lot more or a lot less arguments)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Married male: Change in wife's work outside the home (beginning work, ceasing work, changing to a new job, etc.)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Married female: Change in husband's work outside the home (loss of job, beginning new job, retirement, etc.)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐

Q14.
Continued (5/6)

	Not applicable	0 to 6 months	7 to 12 months	Extremely negative	Moderately negative	Somewhat negative	No impact	Slightly positive	Moderately positive	Extremely positive
Major change in usual type and/or amount of recreation	☐	✔	☐	✔	☐	☐	☐	☐	☐	☐
Borrowing more than £10,000 (buying home, business, etc.)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Borrowing less than £10,000 (buying car, TV, getting school loan, etc.)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Being fired from job	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Male: Wife/girlfriend having abortion	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Female: Having abortion	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐
Major personal illness or injury	☐	✔	☐	✔	☐	☐	☐	☐	☐	☐
Major change in social activities, e.g. parties, movies, visiting (increased or decreased participation)	☐	✔	☐	✔	☐	☐	☐	☐	☐	☐
Major change in living conditions of family (building new home, remodelling, deterioration of home, neighbourhood, etc.)	✔	☐	☐	☐	☐	☐	☐	☐	☐	☐

	Not applicable	0 to 6 months	7 to 12 months	Extremely negative	Moderately negative	Somewhat negative	No impact	Slightly positive	Moderately positive	Extremely positive
Divorce	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐

Q15.
Continued (6/6)

	Not applicable	0 to 6 months	7 to 12 months	Extremely negative	Moderately negative	Somewhat negative	No impact	Slightly positive	Moderately positive	Extremely positive
Serious injury or illness of close friend	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐
Retirement from work	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐
Son or daughter leaving home (due to marriage, college, etc.)	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐
Ending of formal schooling	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐
Separation from spouse (due to work, travel, etc.)	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐
Engagement	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐
Breaking up with boyfriend/girlfriend	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐
Leaving home for the first time	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐
Reconciliation with boyfriend/girlfriend	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐

Q16. Have you had any other experiences that happened in the past year and that had an impact on your life?

◉ Yes
◯ No

Q17. Please list up to three other recent experiences which have had an impact on your life and were not listed.

> Passed qualifying exams to become solicitor in England and Wales. Not job related. Just did it.

Q18.
Please rate the experiences you have listed.

1. indicate the time period during which you have experienced each event

2. indicate the extent to which you viewed the event as having either a positive or negative impact on your life at the time the event occurred.

	0 to 6 months	7 to 12 months	Extremely negative	Moderately negative	Somewhat negative	No impact	Slightly positive	Moderately positive	Extremely positive
Experience 1	☑	☐	☐	☐	☐	☐	☐	☐	☑
Experience 2	☐	☐	☐	☐	☐	☐	☐	☐	☐
Experience 3	☐	☐	☐	☐	☐	☐	☐	☐	☐

Appendix E

COGNITIVE REAPPRAISAL AND WELL-BEING

Daily Outcome Measures

The Positive and Negative Affect Scale (Watson, Clark, & Tellegen, 1988)

Q7.
This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you feel this way <u>right now, that is, at the present moment</u>. Please indicate your responses to all items.

	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
Interested	○	○	○	●	○
Distressed	○	●	○	○	○
Excited	○	○	○	●	○
Upset	○	●	○	○	○
Strong	●	○	○	○	○
Guilty	○	○	●	○	○
Scared	●	○	○	○	○
Hostile	●	○	○	○	○
Enthusiastic	○	●	○	○	○
Proud	●	○	○	○	○
Irritable	○	○	●	○	○
Alert	○	○	●	○	○
Ashamed	●	○	○	○	○
Inspired	●	○	○	○	○
Nervous	●	○	○	○	○
Determined	○	●	○	○	○
Attentive	○	○	○	●	○
Jittery	●	○	○	○	○
Active	○	●	○	○	○
Afraid	●	○	○	○	○

Daily Stress and Frequency of CR Use Measures

Q8. Over the past 24 hours, how stressed did you feel overall?

Not at all stressed	2	3	4	5	6	7	8	9	Extremely stressed
○	○	○	○	○	○	○	○	○	●

Q9. Over the past 24 hours, how often did you use cognitive reappraisal (assigning a more positive meaning to a stressful event)?

Not at all	2	3	4	5	6	7	8	9	All the time
○	○	●	○	○	○	○	○	○	○

COGNITIVE REAPPRAISAL AND WELL-BEING

Appendix F

Pre and Post-Study Outcome Measures

The Patient Health Questionnaire (Kroenke, Spitzer, & Williams, 2001)

The questions in this scale ask you about various difficulties you may have experienced in the last week. Please indicate your responses to all items.

Over the last week, how often have you been bothered by any of the following problems?

	Not at all	Several days	More than half the days	Nearly every day
Little interest or pleasure in doing things	○	○	○	○
Feeling down, depressed, or hopeless	○	○	○	○
Trouble falling or staying asleep, or sleeping too much	○	○	○	○
Feeling tired or having little energy	○	○	○	○
Poor appetite or overeating	○	○	○	○
Feeling bad about yourself - or that you are a failure or have let yourself or your family down	○	○	○	○
Trouble concentrating on things, such as reading the newspaper or watching television	○	○	○	○
Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual	○	○	○	○
Thoughts that you would be better off dead or of hurting yourself in some way	○	○	○	○

COGNITIVE REAPPRAISAL AND WELL-BEING

The Perceived Stress Scale (Cohen, Kamarck, & Merhelstein, 1994)

The questions in this scale ask you about your feelings and thoughts during the last month. In each case, you will be asked to indicate by marking how often you felt or thought a certain way. Please indicate your responses to all items.

	Never	Almost Never	Sometimes	Fairly Often	Very Often
In the last month, how often have you been upset because of something that happened unexpectedly?	○	○	○	○	○
In the last month, how often have you felt that you were unable to control the important things in your life?	○	○	○	○	○
In the last month, how often have you felt nervous or "stressed"?	○	○	○	○	○
In the last month, how often have you felt confident about your ability to handle you personal problems?	○	○	○	○	○
In the last month, how often have you felt that things were going your way?	○	○	○	○	○
In the last month, how often have you found that you could not cope with all the things that you had to do?	○	○	○	○	○
In the last month, how often have you been able to control irritations in your life?	○	○	○	○	○
In the last month, how often have you felt that you were on top of things?	○	○	○	○	○
In the last month, how often have you been angered because of things that were outside of your control?	○	○	○	○	○
In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?	○	○	○	○	○

Cognitive Reappraisal Subscale of the Emotion Regulation Questionnaire

(Gross & John, 2003)

We would like to ask you some questions about your emotional life, in particular, how you control (that is, regulate and manage) your emotions. Please indicate the extent to which you agree or disagree with the six statements below.

	Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
When I want to feel more positive emotion (such as joy or amusement), *I change what I'm thinking about.*	○	○	○	○	○	○	○
When I want to feel less negative emotion (such as sadness or anger), *I change what I'm thinking about.*	○	○	○	○	○	○	○
When I'm faced with a stressful situation, I make myself *think about it* in a way that helps me stay calm.	○	○	○	○	○	○	○
When I want to feel more *positive* emotion, I *change the way I'm thinking* about the situation.	○	○	○	○	○	○	○
I control my emotions by *changing the way* I think about the situation I'm in.	○	○	○	○	○	○	○
When I want to feel less *negative* emotion, I *change the way I'm thinking* about the situation.	○	○	○	○	○	○	○

Appendix G

Intervention 1: Cognitive Reappraisal Task

About Cognitive Reappraisal

Cognitive reappraisal is a technique that can help us to change the way we react to a stressful event by actively changing the meaning we make of this event.

'Cognitive' refers to the use of thinking and 'reappraisal' means we are re-appraising or re-evaluating the meaning of something.

Experiencing stressful life events can have a serious impact on our physical and mental well-being. However, research shows that we do have some control over how we are going to react to such experiences. By changing the meaning we make of something stressful that happened to us, we can change how we feel afterwards. This is not always easy, but practice makes perfect. The example on the next page can give you a good idea of how cognitive reappraisal can be done.

If you have questions about cognitive reappraisal, please ask the researcher.

Example Use of Cognitive Reappraisal

Below are two paragraphs. The first one gives an example of a stressful event described by the person who experienced it. In the second paragraph this person used cognitive reappraisal to gain a new perspective on the event.

Description of a stressful event:

Yesterday I was late for my doctor's appointment. The school run was taking forever and my son did not want to leave the car. He's 6 and still finds it difficult to spend all day in school. When I got to the surgery, I was told I have to reschedule, because I was more than 15 minutes late. I had an argument with the receptionist who always pushes my buttons and eventually managed to see the doctor. When I was in the clinic room I became tearful when I was talking about my physical health problems. It was just too much. I felt really embarrassed and wanted to leave. That morning was just terrible!

Description of the event using cognitive reappraisal:

Yesterday I was late for my doctor's appointment. The school run often takes a long time, but it's nothing that hadn't happened before. Traffic that day was pretty heavy – no wonder, other parents would have to drop their children off too. My son is only six and he often wants to stay in the car with me for a bit longer when we get to the school gate. I do love our car conversations before school, although they can make me late. I know these conversations are important to him, because he's not distracted by anything else. I was late for my doctor's appointment afterwards, but managed to negotiate with the receptionist, despite being a bit edgy and irritable that morning. I really don't like how short she can be with people, but I can see it's a busy job! The doctor was really supportive when I teared up in her office. I got what I needed and although I felt embarrassed about becoming emotional in front of her, I feel like we have a warmer relationship now. Her daughter is in my son's school. She's adorable. It was a stressful morning and I felt exhausted, but I managed to get everything done!

Brilliant, let's practice this using a real situation.

Try to think of a stressful event that happened in the past 48 hours.

It may have been something relatively small or fairly significant. See a few examples below:

- A broken carrier bag on your way from work
- Argument with a member of family
- Delayed / missed train
- A passer-by in town was aggressive
- There were too many people in your shop and you found it stressful

We understand that, on their own, many traumatic events can be seen as relatively small and "insignificant", but their power and impact they can have is due to the context in which they take place. In other words, often it is not the event itself that makes us feel difficult emotions - it is everything else that is happening in our lives which causes us stress and the event then functions as a trigger, the last drop that pushes us over the edge and makes us feel bad.

Please provide a description of a difficult event that you experienced in the past 48 hours. At this point, **do not** use cognitive reappraisal. We just want a description of what happened including any thoughts you may have had. Your main focus should be on the event, but feel free to describe the context of the event too.

Thank you.

Now try to re-write the descpriation using cognitive reappraisal. Remember, it can be helpful to try use other perspectives when reappraising what happened, although it is better to be realistic. **Do not change** *what happened*, change *the way you view* **the event and its context.**

There are no right or wrong answers as the situation you described can be re-written in many different ways.

You can scroll up to look at your original description.

COGNITIVE REAPPRAISAL AND WELL-BEING

Appendix H

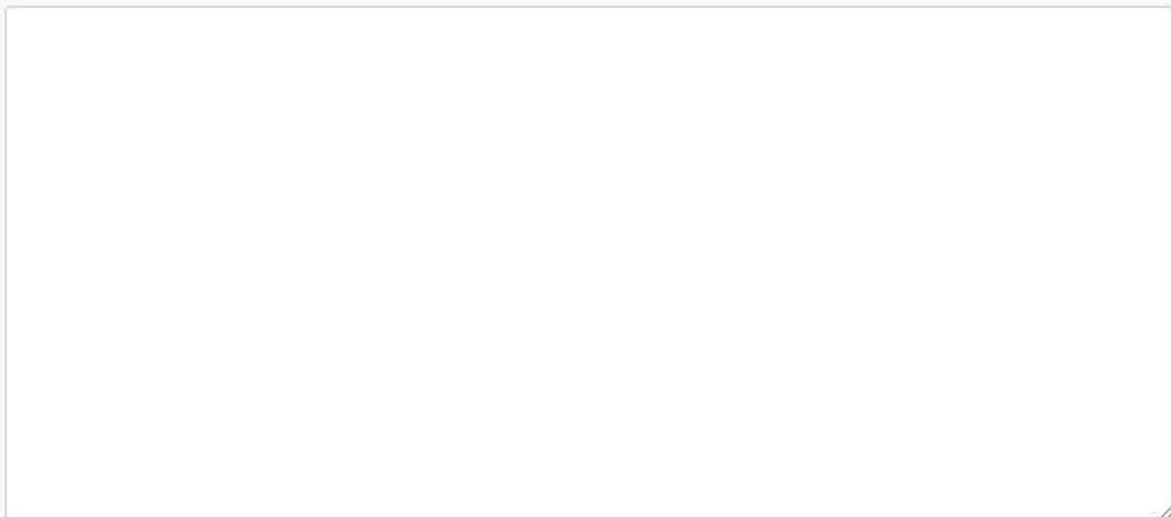Intervention 2: Cognitive Reappraisal Helping Task

Day 1 of CR Helping Task

The next cognitive reappraisal exercise will be slightly different to the ones you have been completing until now. We would like to ask you to help another participant in this study to reappraise their stressful experience. As you may have noticed, sometimes it is hard to gain a new perspective on your own experience.

Using your cognitive reappraisal skills that you have learnt, please rewrite the description of a stressful event in the space provided below the text. Do not include any identifiable information (for example your name, address, contact details).

TEXT:

It was Wednesday morning and I woke up feeling down. I remember talking to my husband about not wanting to go to work, but he was not really helpful. Anyway I got to work (I'm a nurse) and realised my shift had been cancelled. I remember feeling quite positive for a minute coz it meant I could go home but then realised I won't get paid of course! I felt irritated on the way home coz I thought they should have informed me earlier. On my way home I don't know how it happened, but I nearly ran over some couple in my town. I had so many thoughts in my head and just stopped concentrating for a second. I felt so guilty when the man approached me and I felt I was turning red. He asked me if I was ok. It caught me by surprise coz I honestly thought he was going to shout at me. I quickly apologised and drove away as I was feeling my throat closing. When I got home I honestly felt like crying, but nothing came out.

Day 2 of CR Helping Task

My mate Henry came over for dinner last night. He's a great friend of mine but sometimes he just pushes my buttons!! Henry loves my PlayStation and always wants to game after we eat. So we did the same thing this time, but I was getting sort of bored and stopped caring which I think annoyed him. There was a moment when I said something like "Yeah well you don't know when to stop" and he just completely flipped!! Called me all sorts of names and then left I was in shock. Called him later but he wouldn't answer. I just didn't know what to do when he got angry. I only meant to say he spends a lot of his time playing computer games and stuff and I know he has problems but I do think it can stop him from going out and meeting people and having a job. Anyhow I couldn't sleep all night so I was a wreck this morning at work which didn't help.

Day 3 of CR Helping Task

One of my work colleagues asked me to come to her birthday party next week. I was really pleased, but realised I have to go to my mum's and help her on that day as her carer will be on holiday. I told my colleague this and apologised, but she just did not seem to understand why I cannot pay someone to cover for the carer. I tried explaining that I don't get to spend much time with my mum and so not only I'll save money, but also will be able to have a chat with her. She's almost 90 and has been poorly for a while so I am really worried. My colleague seemed really disappointed and I had the feeling that she avoided me for the rest of the day. I felt really bad because I could see she really wanted people to come to her party. So now I don't really know what to do and it feels like whatever I do I can't win.

Day 4 of CR Helping Task

Last night I went to my boyfriend's. When I arrived I found out one of his friends was there so I greeted them both and sat with the friend while my boyfriend was sorting stuff out in the kitchen. I'd met Mike before so it was pretty cool to catch up with him. After about 20 minutes I got up and went to check up on my boyfriend to see if I could help with anything. He seemed really grumpy and irritable and told me I could have come earlier. I really didn't know what to make of it, because he loves cooking and we sort of take turns who cooks for whom, so it really surprised me what he'd said! Anyway he was proper sulking so I sort of laughed about it and apologised, although I really thought he was overreacting. It made it so much worse! He thought I wasn't taking him seriously and nearly kicked me out of the house. Mike kind of noticed something was wrong so he left. Really awkward. My boyfriend did not want to talk about it so we got sort of stuck. Eventually I stopped trying and left. Ended up having a pretty crap evening on my own. Maybe I should have been more sensitive, it would not have been the first time I got things wrong.

Day 5 of CR Helping Task

There are eight people living in our house and sometimes it gets so messy. I mean in terms of relationships. But cleanliness is also a problem come to think of it. My housemate who lives in the bedroom next to mine played music last night and like I love his taste but I could just hear the base even when it was super quiet. Told him twice to turn it down but he kept telling me it's not even 10 but I was just so knackered pulled a few all nighters this week coz with work I have to do my language course in the evenings and sometimes I just fall asleep and then wake up after midnight and realise I have to study. My mum keeps telling me to go to sleep early lol. Yeah so basically my housemate was a pillock and I couldn't sleep and was angry with him and then started thinking about everything that's wrong with our house and it took this weird turn I basically started planning how I am going to move out and like not even tell them. I was sooooo angry ended up doing some work I know I should have been more assertive and like go punch him in the face or something but I just couldn't even move because of how angry I was. It paralyses me I know most people snap but I just freeze and die on the inside it's the worst feeling not sure if it happens to other people. Guess I've got issues? That was a bad night.

COGNITIVE REAPPRAISAL AND WELL-BEING

Appendix I

Informed Consent Form

UNIVERSITY OF
EXETER

CONSENT FORM

Title of Project: **Using Cognitive Reappraisal and Helping Behaviour to Improve Well-being: Single-Case Design Study**

Name of Researcher: **Tomas Jelinek**

Name of Supervisors: Dr. Pia Pechtel, Dr. Nick Moberly

Please initial box

1. I confirm that I have read the information sheet for the above project. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without my legal rights being affected. There will be no consequences to my withdrawal or discontinuation of the study.

3. I understand that relevant sections of the data collected during the study may be looked at by members of the research team, and individuals from the University of Exeter where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.

4. I understand that taking part involves anonymised data responses will be stored for a period of up to *5* years.

5. I understand that taking part in this study means that anonymised data responses can be used in published reports.

6. I understand that I will be reimbursed £40 upon the completion of the study. Should I withdraw or become unable to complete the full length of the study (21 days), I will be paid on a pro rata basis.

7. I agree for my anonymised questionnaire responses to be shared with other researcher for use in future research projects.

8. I agree that my contact details can be kept securely as I would like to be contacted by other researchers about future research projects.

9. I agree to be contacted via telephone should my answers to risk related questions in online questionnaires indicate I am at risk of harming myself, or others. I understand I will be contacted within 48 hours (including weekends) from the point of indicating risk.

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

_____   _____   _____

Name of Participant   Date   Signature

_____   _____   _____

Name of researcher
taking consent   Date   Signature

Appendix J

The Mood Disorder Centre Protocol for Assessing and Reporting Risk

MOOD DISORDERS CENTRE
PROTOCOL FOR ASSESSING AND REPORTING RISK

The following principles and procedures govern risk assessment and reporting in the Mood Disorders Centre (MDC). The MDC does not does manage risk.

General principles

MDC clinical academic faculty are responsible for risk assessment in their research programmes. This includes ensuring that staff, students and interns working with them receive adequate induction, training and supervision.

Many of the research projects in the MDC will include supplementary and more detailed protocols for risk assessment.

The AccEPT Clinic Directors (for new assessments) and clinic therapists (for patients in therapy) are responsible for risk assessment in the AccEPT clinic. The detailed Devon Primary Care Trust guidelines for risk assessment are used in AccEPT.

General procedures

Background training materials are available on the shared directory.

Whenever any significant risk is identified a risk assessment should be completed and (counter-) signed by the responsible member of staff. If at all possible this should be done at the time of the assessment, or as soon afterwards as possible.

Any significant, but not imminent risk should be reported to the person's GP and, if appropriate, other health care professionals, as soon as is reasonably possible.

Any imminent risk should lead to the immediate involvement of the appropriate emergency health services.

When clinical academic staff are away from the Centre they should ensure appropriate cover is arranged for any risk issues that might arise in their absence.

MDC Risk Protocol – version 11/03/09

Exploring Risk in Research Interviews

THOUGHTS

"I see that you've said / you mentioned that…… … These are thoughts / feelings that people suffering from depression often have, but it's important to make sure you are receiving the right kind of support. So if it's OK, I would now like to ask you some more questions that will explore these feelings in a little more depth."

PLANS

1 Do you know how you would kill yourself?   Yes / No
If **yes** – details

---

2 Have you made any actual plans to end your life?   Yes / No
If **yes** – details

ACTIONS

3 Have you made any actual preparations to kill yourself?   Yes / No
If **yes** – details

---

4 Have you ever attempted suicide in the past?   Yes / No
If **yes** – details

PREVENTION

5 Is there anything stopping you killing or harming yourself
at the moment?   Yes / No
If **yes** – details

---

6 Do you feel that there is any immediate danger that you
will harm or kill yourself?   Yes / No
Details:

See risk table overleaf for appropriate actions.

Researcher Risk Protocol

To be used following any indication of risk from questionnaire items, responses to interview questions or any other sources.

Look at answers from the sheet to determine the level of risk, A B or C:

Actions by Researcher	***Tell Participant***
All telephone follow-up interviews	*As part of our standard procedure we usually ask people their current whereabouts because very occasionally we may be concerned about people's safety and need to get some assistance for them straight away.*

All answers 'no' apart from Q5 'yes':

↓

A

I can see that things have been very difficult for you, but it seems to me these thoughts about death are not ones you would act on – would this be how you see things? (if they say yes) I would advise you to make an appointment to see your GP to talk about these feelings

'Yes' for any **one** of Qs 1-4; plus 'yes' for Q5 and 'no' for Q6

↓

B

Things seem to be very hard for you right now and I think it would help if you were to speak to your GP about these feelings. I would also advise you to make an appointment to see your GP to talk about these feelings. In these circumstances we usually write to people's GP as well to tell them that they have been seen by us and have been having some troubling thoughts. Would you happy for me to let the Student Health Centre know?

Scoring 'no' to Q5 or 'yes' to Q6

↓

C Actively Suicidal

I am very concerned about your safety at this moment, I am not a therapist but I would like you to talk to one right now. I am going to make some telephone calls now to arrange for someone to come and talk to you.

Participant needs immediate help – **do not leave them alone**. Follow your trial's chain of supervisory clinical contact and enact immediate risk procedure. (For follow-up telephone interviews researcher to alert Student Health Centre, emergency services or crisis team whilst maintaining contact with participant or immediately get help from supervisor or named psychologist from MDC to contact assistance).

Appendix K

Debriefing Information

Thank you for completing the questionnaires. Please alert the researcher.

Debrief

This study aimed to achieve a couple of things. We wished to see whether a repeated use of cognitive reappraisal for daily stressful events could help you to feel better. More specifically, we were interested in emotional well-being and depressive symptoms.

As part of the study you may have also been asked to use cognitive reappraisal to reappraise other participants' descriptions of stressful event. We see this as helping behaviour and the study aim was also to find out, whether you feel better when you have helped a fellow participant compared to when you have used cognitive reappraisal only for yourself. Research suggests that helping through peer support may be beneficial not only to the recepient, but also to the helper.

Some participants in this study were using cognitive reappraisal only for themselves and some were asked to also help others. The scenarios we asked you to reappraise to help others were not written by any of the participants in this study. However, the scenarios were checked by the Lived Experience Group and we were assured they are good representations of what may happen to people on bad days when they feel low in mood.

If you have questions about this study, please speak with the researcher.

Finally, if you are interested in learning about the results of this study, please inform the researcher.

Thank you. This is the end of the study.

You can collect payment and speak with the researcher should you have any questions.

Please click NEXT to finish.

Appendix L

Ethical Approval

CLES – Psychology
Psychology
College of Life and Environmental Sciences
University of Exeter
Washington Singer Building
Perry Road
Exeter
EX4 4QG
Web: www.exeter.ac.uk

CLES – Psychology Ethics Committee

Dear Tomas Jelinek

Ethics application - eCLESPsy000606
Using Cognitive Reappraisal and Helping Behaviour to Improve Well-being: Single-Case Design Study

Your project has been reviewed by the CLES – Psychology Ethics Committee and has received a Favourable with conditions opinion.

The Committee has made the following comments about your application:
This was a very thorough and thoughtful application.   Please respond to the following queries:
(1) How long will it take to contact high risk participants (if risk is discovered via online scores)? 24, 48
hours? What about on weekends? This should be included in the consent form.   (2) if the research
ers become aware of risk - e.g., 2 or 3 on the Phq-9, then it is not appropriate to only exclude and
tell participant to follow-up with GP. If the person is at high risk of suicide, then the researche
- Please view your application at https://eethics.exeter.ac.uk/CLESPsy/ to see comments in full.
If you have received a Favourable with conditions, Provisional or unfavourable outcome you are required to re-submit for
full review and/or confirm that committee comments have been addressed before you begin your research.

If you have any further queries, please contact your Ethics Officer.

Yours sincerely

Date: 02/10/2018

CLES – Psychology Ethics Committee

Appendix M

Detailed Description and Justification of all Methods of Data Analysis Used   in the Study

Visual analysis uses a graphic display of data and draws conclusions regarding the reliability and consistency of intervention effects by visual inspection of data patterns within and between individual participants (Lane & Gast, 2014). Therefore, it has long been considered the most sensitive and appropriate method to detect intervention effects in SCED studies (Parsonson & Baer, 1986), although it is now widely accepted that adding a statistical approach improves the quality of SCED data analysis by removing observer bias from visual judgements (Jones, Weinrott, & Vaught, 1978) and potentially detecting smaller effects that visual analysis may disregard (Kazdin, 1982). Several guidelines for systematically analysing data exist (Kratochwill et al., 2010; Lane & Gast, 2014), yet decision-making criteria can vary. In keeping with the SCRIBE guidance (Tate et al., 2016), this study uses visual analysis features of (1) stability, (2) level, and (3) trend to evaluate individual phases as well as (4) changes in level, (5) trend, and (6) overlap of data to examine changes in outcomes between phases.

Within phases evaluation*.*

Stability. Stability of data within phases was assessed using the stability criterion. Following Lane and Gast's (2013) recommendation, data within a phase may be considered stable if 80% of the data fall within +/- 25% of the median. Instability of baseline data effects the reliability with which causal inferences about changes between phases can be made (Morley, 2017).

Level. Level of data within phases was assessed using the median, a robust estimator of central tendency that is represented by the number in the middle of the data when rank ordered. Median is considered an appropriate method of level estimation in small data sets where the mean can be affected by extreme scores (Morley, 2017).

Trend. The split-middle method was employed to assess the linear trend of data. This method finds two data points that best represent 'average' in each half of a phase. This is done by identifying the median data value and time point in each half, yielding two coordinates that can be plotted and connected with a straight line. Compared to a best-fit line, for example, the split-middle method is resistant to the influence of outliers and is particularly valuable in small data sets (Morley, 2017).

Between phases evaluation.

Changes in level. Changes in level between phases were assessed by comparing median values of consecutive phases.

Changes in trend. Changes in trend between phases were assessed by visual inspection of trend lines in consecutive phases.

Overlap of data. Overlap of data between phases was explored using the non-overlap of all pairs index (NAP; Parker & Vannest, 2009). The NAP is computed by comparing each data point in the phase where change is expected with each data point in the baseline. A score of 0 is given where there is no improvement and data overlap between phases. A score of 1 is given where improvement was observed and data do not overlap. The total number of

improvements is divided by the total number of comparisons yielding a score that represents the proportion of all possible comparisons of pairs where desired change in data was observed. Scores closer to one indicate a small number of overlapping pairs and therefore a larger effect of an intervention (Jamieson, Cullen, McGee-Lennon, Brewster, & Evans, 2014). The NAP was calculated using a web-based calculator (Vannest, Parker, Gonen, & Adiguzel, 2016). In keeping with conventions for NAP effect sizes, values of .50 - .65 suggested a weak effect of the intervention, .66 - .92 a medium effect and .93 – 1.00 a strong effect (Parker & Vannest, 2009). Because participants' Frequency of CR Use and CR Ability continued to improve across phases, data from intervention and follow-up were combined into one phase and compared with data in baseline.

   Randomisation test. Randomisation tests are used to evaluate the statistical significance of SCED outcomes by testing the probability of a particular set of observations occurring given all the possible ways in which the data can be arranged (Onghena & Edgington, 2005). Randomisation tests are particularly useful for SCED analysis as they are non-parametric and do not make assumptions about the nature of the error structure in the data or sampling distributions of parameters (Morley, 2017). In addition, randomisation tests are frequently used when conclusions made from visual analysis are not clear (Bulté & Onghena, 2013). Participants must be randomized to moment of phase change a priori to control against threats to validity such as history or maturation. For b possible intervention start points and N participants, there are b to the power N (b^N) possible randomization start points for the intervention

(Morley, 2017). Higher number of data points provide greater statistical power. In this study, when assessing effects of the intervention, for three possible moments of phase change with twelve participants, this yields 531,441 (3^{12}) possible start points, and a minimum potential *p* value of 0.00000188 (1/531441). Heyvaert and colleagues (2017) suggested that a minimum of 20 measurement times and 5 participants is required to achieve 80% power for a specified significance test with alpha = .05. Consequently, this study has enough power to detect statistically significant change across phases at *p* = .05. The randomisation test was carried out in R software utilising a Monte-Carlo simulation (Bulté & Onghena, 2009). Missing data points were managed by calculating the median for the nearest five MTs within the same phase in the relevant subscale.

Evaluation of changes from pre- to post-intervention.

Reliable Change Index. The Reliable Change Index (RCI) is used to determine if clinically meaningful change occurred as a result of clinical interventions (Jacobson & Truax, 1991). RCI is defined as the change in an individual's score divided by the standard error of the difference for the measure that is used. Therefore, it provides an estimate of relative change and controls for the measure's reliability (Duff, 2012). In this study, RCI was used to assess change in pre and post measures (PHQ-9, PSS) and to examine such differences in the two intervention conditions (intervention 1 and 2).

Appendix N

Results of Repeated Measures ANOVA Assessing Differences in Frequency of CR Use and CR Ability between Phases

A repeated measures ANOVA showed that participants were using CR significantly more frequently during intervention and follow-up compared to baseline [$F(1, 11) = 8.38$, $p = .015$]. Participants continued to use CR in follow-up. The difference in Frequency of CR Use between intervention and follow-up was not significant [$F(1, 11) = 0.62$, $p = .81$]. These results suggest that, as a group, the intervention significantly increased Frequency of CR Use and that participants continued to use CR with similar frequency even after the intervention was discontinued. Frequency of CR Use did not increase for participants 1 and 7.

The intervention also improved participants' CR Ability. Differences in CR Ability were statistically significant between baseline scores and follow-up [$F(1, 11) = 17.84$, $p = .001$], but were not significant when phases B and C were compared to baseline in combination [$F(1, 11) = 4.33$, $p = .06$]. Thus the effect of training in CR emerged at the follow-up stage, rather than during active training in the intervention itself. CR Ability did not improve for participants 1 and 7.

COGNITIVE REAPPRAISAL AND WELL-BEING

Appendix O

Within Phase Visual Analysis of Daily Outcome Measures

Positive Affect

Stability. Evaluation of the baseline phase using the stability criterion indicated data were stable in three participants (4, 5, and 6) in group 1 and five participants (7, 8, 9, 10, and 12) in group 2. Within the intervention phase, stability criterion was met in four participants (1, 2, 3, and 6) in group 1 and three participants (8, 11, and 12) in group 2. In follow-up, data were stable in four participants (1, 2, 3, and 6) in group 1 and four participants (8, 9, 11, and 12) in group 2.

Level. Compared to a normative sample, evaluation of level within baseline using the median indicated baseline PA was extremely low in two participants (10 and 11) in group 2, relatively low in five participants (1, 3, 4, 5, and 6) in group 1 and two participants (7 and 8) in group 2, and about average in one participant (2) in group 1 and three participants (8, 9, and 12) in group 2. Within the intervention phase, median PA was extremely low in one participant (1) in group 1 and two participants (10 and 11) in group 2, relatively low in three participants (4, 5, and 6) in group 1 and three participants (7, 8, and 9) in group 2, about average in one participant (3) in group 1 and one participant (12) in group 2, and relatively high in one participant (2) in group 1. Within follow-up, PA was extremely low in in two participants (1 and 5) in group 1 and two participants (10 and 11) in group 2, relatively low in three participants (4, 5, and 6) in group 1 and three participants (7, 8, and 12) in group 2, about average in

two participant (3) in group 1 and one participant (9) in group 2, and relatively high in one participant (2) in group 1.

Trend. Evaluation of trend within baseline using the split-middle method indicated a decreasing, contra-therapeutic trend in four participants (1, 4, 5, and 6) in group 1 and five participants (7, 8, 10, 11, and 12) in group 2. PA data showed an increasing, therapeutic trend in participants 2 and 3. Within the intervention phase, five participants (2, 3, 4, 5, and 6) in group 1 and three participants (9, 10, and 12) in group 2 showed a decreasing, contra-therapeutic trend in PA data. Only three participants (7, 8, and 11; group 2) showed a slightly increasing, therapeutic trend in PA. In follow-up, three participants (3, 4, and 6) in group 1 and one participant (11) in group 2 showed a decreasing, contra-therapeutic trend. Five participants (7, 8, 9, 10, and 12; group 2) showed an increasing, therapeutic trend in PA data.

Negative Affect

Stability. Evaluation of the baseline phase using indicated data were stable in three participants (3, 4, and 5) in group 1 and three participants (9, 11, and 12) in group 2. Within the intervention and follow-up phases, data were stable in all six participants in group 1 and five participants (7, 8, 9, 11, and 12) in group 2.

Level. Compared to a normative sample, evaluation of level within baseline using the median indicated baseline NA was extremely low in three participants (2, 3, and 4) in group 1 and four participants (8, 9, 11, and 12) in

group 2, relatively low in one participant (5) in group 1, about average in one participant (1) in group 1 and one participant (7) in group 2 and relatively high in one participant 6) in group 1 and one participant (10) in group 2. Within the intervention phase, median NA was extremely low in five participants (1, 2, 3, 4, and 5) in group 1 and four participants (8, 9, 11, and 12) in group 2, about average in one participant (6) in group 1 and one participant (7) in group 2, and relatively high in one participant (10) in group 2. Within follow-up, NA was extremely low in five participants (1, 2, 3, 4, and 5) in group 1 and four participants (8, 9, 11, and 12) in group 2, about average in one participant (6) in group 1 and one participant (7) in group 2.

Trend. Evaluation of trend within baseline using the split-middle method indicated a decreasing, therapeutic trend in two participants (1 and 2) in group 1 and four participants (8, 9, 11, and 12) in group 2. NA data showed an increasing, contra-therapeutic trend in three participants (4, 5, and 6) in group 1 and one participant (7) in group 2. Within the intervention phase, one participant (5) in group 1 and one participant (9) in group 2 showed an increasing, contra-therapeutic trend in NA. One participant (6) in group 1 and four participants (7, 8, 10, and 11) in group 2 showed a decreasing, therapeutic trend in NA. In follow-up, two participants (2 and 4) in group 1 and four participants (7, 8, 10, and 12) in group 2 showed an increasing, contra-therapeutic trend. Only two participants (1 and 6; group 1) showed a decreasing, therapeutic trend in NA.

Daily Stress

Stability. Evaluation of the baseline phase using the stability criterion indicated data were stable in three participants (1, 5, and 6) in group 1 and three participants (8, 9, and 10) in group 2. Within the intervention phase, data were stable in one participant (5) in group 1 and two participants (8 and 9) in group 2. In follow-up, DS scores were stable in three participants (1, 4, and 6) in group 1 and two participants (8 and 12) in group 2.

Level. Evaluation of level within baseline using the median indicated DS of 6 and above was present in four participants (1, 4, 5, and 6) in group 1 and four participants (7, 9, 10, and 11) in group 2. Within the intervention phase, median DS of 6 and above was present in three participants (4, 5, and 6) in group 1 and two participants (8 and 10) in group 2. In follow-up, DS of 6 and above was observed in three participants (1, 5, and 6) in group 1 and one participant (10) in group 2.

Trend. Evaluation of trend within baseline using the split-middle method indicated a decreasing, therapeutic trend in two participants (2 and 3) in group 1 and three participants (10, 11, and 12) in group 2. DS data showed an increasing, contra-therapeutic trend in thee participants (4, 5, and 6) in group 1 and one participant (7) in group 2. Within the intervention phase, two participants (4 and 6) in group 1 and two participants (8 and 10) in group 2 showed an increasing, contra-therapeutic trend in DS. Two participants (3 and 5) in group 1 and two participants (7 and 9) in group 2 showed a decreasing,

therapeutic trend in DS. In follow-up, one participant (6) in group 1 and four participants (7, 9, 10, and 11) in group 2 showed an increasing, contra-therapeutic trend. Only participants (1 and 5; group 1) showed a decreasing, therapeutic trend in DS.

Appendix P

Dissemination Statement

Results of the study will be disseminated in several ways. Firstly, all participants who took part in the study were asked whether they would like to be informed of the results of the study. All twelve individuals expressed interest and will be sent a summary of the results via e-mail or postal mail (depending on indicated preference) in June 2019.

The study will also be disseminated via an oral presentation at Exeter University in June 2019. This has been organised to share the study findings with colleagues and other professionals.

Finally, the revised manuscript of the empirical study will be submitted for publication to a peer-reviewed journal Emotion (see Appendix Q for manuscript submission guidelines).

Appendix Q

Manuscript Submission Guidelines for *Emotion*

Length of Manuscripts
Manuscripts for *Emotion* will typically range from 10 to 40 double-spaced manuscript pages, including the cover page, abstract, text, references, tables and figures, with margins of at least 1 inch on all sides and a standard font (e.g., Times New Roman) of 12 points (no smaller).
For manuscripts that exceed 40 pages, authors must justify the extended length in their cover letter (e.g., the inclusion of multiple studies). The entire manuscript must be double spaced.

Title of Manuscript
The title of a manuscript should be accurate, fully explanatory, and preferably no longer than 12 words.

Cover Letter
The cover letter accompanying the manuscript submission must include all authors' names and affiliations to avoid potential conflicts of interest in the review process. Addresses and phone numbers, as well as email addresses and fax numbers, if available, should be provided for all authors for possible use by the editorial office and later by the production office.

Evaluation Criteria
Manuscripts submitted to *Emotion* will be evaluated based on the extent to which they make theoretical and empirical contributions that advance an understanding of emotional processes. Articles will be evaluated in terms of the overall theoretical and empirical contribution of the work, including the soundness of the research methodology, appropriateness of the statistical analyses, and accuracy in interpreting the findings.

Types of Articles
Most of the articles published in *Emotion* will be reports of original empirical research, but other types of articles will be considered.
These include:
- Articles that present or discuss theoretical formulations of emotion and related affective phenomena that evaluate competing theoretical perspectives, or that offer innovative commentary or analysis on timely topics of inquiry.
- Comprehensive reviews of the empirical literature in an area of study that contain a meta-analysis and/or present novel theoretical or methodological perspectives.
- Comments on articles published in the journal.
- Case studies from either a clinical setting or a laboratory that raise or illustrate important questions that go beyond the single case and have heuristic value.

Brief Reports

Emotion also publishes brief reports. Manuscripts submitted as Brief Reports should not exceed 2,500 words, exclusive of references and figure captions. There should be no more than 2 figures or tables and no more than 30 references.

Commentaries

Emotion will consider commentaries on articles previously published in this journal. Comments will be sent out for external review and must meet a high threshold for publication.

Comments will be evaluated based on how accurately and fairly they represent the target article, the importance of the comment in providing a more comprehensive understanding of the target article, and the clarity, conciseness and tone of the writing. External reviewers will include individuals not associated with the target article as well as the corresponding author of the target article. Two types of commentaries will be considered.

Brief Comment

A Brief Comment is written in response to a single article previously published in *Emotion*. The primary purpose is to provide a meaningful insight, concern, alternative interpretation, clarification, or critical analysis. The comment should provide a richer and more comprehensive understanding of a methodological, conceptual, or interpretation issue that significantly adds to the literature.

Brief Comments should not exceed 265 lines of text including references. This limit does not include the title page, abstract, or author notes.

The title of a Brief Comment should include a subtitle reflecting the actual title and year of publication of the article that engendered the comment. For example — "The Importance of Focusing on External Validity: A Brief Comment on *Testing the Efficacy of Two Differing Types of Stress Management Interventions for the Treatment of Essential Hypertension* (Jones & Smith, 2012)."

Brief Comments must be submitted in a timely manner, no later than 9 months after publication of the original article.

Upon acceptance of a Brief Comment, the author(s) of the original paper would be invited to submit a response; if the authors' response is acceptable, both the Brief Comment and Response would be published together. Such responses to a Brief Comment also should not exceed 265 lines of text including references.

Extended Comment

The purpose of an Extended Comment is similar to that of a Brief Comment (i.e., to provide a meaningful insight, concern, alternative interpretation, clarification, or critical analysis), but it would be written in response to a series of articles previously published in *Emotion* or that involves a more extensive and far-reaching conceptual or methodological issue.

An example might include a comment describing and analyzing the limitations of a particular statistical or methodological procedure used in several studies previously published in *Emotion*; the comment also must include recommendations for addressing such limitations.

Extended comments will be evaluated on the timeliness of the topic and the potential contribution to the scientific literature relevant to the scope of *Emotion*.

This type of article should not exceed approximately one half the length of the original paper (note that 1 journal page equals approximately 3–3.5 manuscript pages). Unless permission from the editor is received, no Extended Comment should exceed 20 manuscript pages inclusive of all references, tables, and figures.

If the Extended Comment is accepted, the author(s) of the original article(s) will be contacted to write a response; and, if the authors' responses are accepted, both the Extended Comment and Response(s) would be published together. This Invited Response should not exceed approximately one half the length of the Extended Comment.

The title of this type of article need not include a subtitle representing the original article(s).

Manuscript Preparation

Prepare manuscripts according to the *Publication Manual of the American Psychological Association* (6[th]edition). Manuscripts may be copyedited for bias-free language (see Chapter 3 of the *Publication Manual*).

Review APA's Journal Manuscript Preparation Guidelines before submitting your article.

Double-space all copy

Other formatting instructions, as well as instructions on preparing tables, figures, references, metrics, and abstracts, appear in the *Manual*. Additional guidance on APA Style is available on the APA Style website.

Below are additional instructions regarding the preparation of display equations, computer code, and tables.

Display Equations

We strongly encourage you to use MathType (third-party software) or Equation Editor 3.0 (built into pre-2007 versions of Word) to construct your equations, rather than the equation support that is built into Word 2007 and Word 2010. Equations composed with the built-in Word 2007/Word 2010 equation support are converted to low-resolution graphics when they enter the production process and must be rekeyed by the typesetter, which may introduce errors.

To construct your equations with MathType or Equation Editor 3.0:

- Go to the Text section of the Insert tab and select Object.
- Select MathType or Equation Editor 3.0 in the drop-down menu.

  If you have an equation that has already been produced using Microsoft Word 2007 or 2010 and you have access to the full version of MathType 6.5 or later, you can convert this equation to MathType by clicking on MathType Insert Equation. Copy the equation from Microsoft Word and paste it into the MathType box. Verify that your equation is correct, click File, and then click Update. Your equation has now been inserted into your Word file as a MathType Equation.

  Use Equation Editor 3.0 or MathType only for equations or for formulas that cannot be produced as Word text using the Times or Symbol font.

Computer Code

Because altering computer code in any way (e.g., indents, line spacing, line breaks, page breaks) during the typesetting process could alter its meaning, we treat computer code differently from the rest of your article in our production process. To that end, we request separate files for computer code.

In Online Supplemental Material
We request that runnable source code be included as supplemental material to the article. For more information, visit Supplementing Your Article With Online Material.

In the Text of the Article
If you would like to include code in the text of your published manuscript, please submit a separate file with your code exactly as you want it to appear, using Courier New font with a type size of 8 points. We will make an image of each segment of code in your article that exceeds 40 characters in length. (Shorter snippets of code that appear in text will be typeset in Courier New and run in with the rest of the text.) If an appendix contains a mix of code and explanatory text, please submit a file that contains the entire appendix, with the code keyed in 8-point Courier New.

Tables
Use Word's Insert Table function when you create tables. Using spaces or tabs in your table will create problems when the table is typeset and may result in errors.

Academic Writing and English Language Editing Services
Authors who feel that their manuscript may benefit from additional academic writing or language editing support prior to submission are encouraged to seek out such services at their host institutions, engage with colleagues and subject matter experts, and/or consider several vendors that offer discounts to APA authors.
Please note that APA does not endorse or take responsibility for the service providers listed. It is strictly a referral service.
Use of such service is not mandatory for publication in an APA journal. Use of one or more of these services does not guarantee selection for peer review, manuscript acceptance, or preference for publication in any APA journal.

Submitting Supplemental Materials
APA can place supplemental materials online, available via the published article in the PsycARTICLES®database. Please see Supplementing Your Article With Online Material for more details.

Abstract and Keywords
All manuscripts must include an abstract containing a maximum of 250 words typed on a separate page. After the abstract, please supply up to five keywords or brief phrases.

References

List references in alphabetical order. Each listed reference should be cited in text, and each text citation should be listed in the References section.
Examples of basic reference formats:
Journal Article:
Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin, 139,* 133–151. http://dx.doi.org/10.1037/a0028566
Authored Book**:**
Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.*Cambridge, MA: MIT Press.
Chapter in an Edited Book:
Gill, M. J., & Sypher, B. D. (2009). Workplace incivility and organizational trust. In P. Lutgen-Sandvik & B. D. Sypher (Eds.), *Destructive organizational communication: Processes, consequences, and constructive ways of organizing* (pp. 53–73). New York, NY: Taylor & Francis.

Figures
Graphics files are welcome if supplied as Tiff or EPS files. Multipanel figures (i.e., figures with parts labeled a, b, c, d, etc.) should be assembled into one file. The minimum line weight for line art is 0.5 point for optimal printing.
For more information about acceptable resolutions, fonts, sizing, and other figure issues, please see the general guidelines.
When possible, please place symbol legends below the figure instead of to the side.
APA offers authors the option to publish their figures online in color without the costs associated with print publication of color figures.
The same caption will appear on both the online (color) and print (black and white) versions. To ensure that the figure can be understood in both formats, authors should add alternative wording (e.g., "the red (dark gray) bars represent") as needed.
For authors who prefer their figures to be published in color both in print and online, original color figures can be printed in color at the editor's and publisher's discretion provided the author agrees to pay:
- $900 for one figure
- An additional $600 for the second figure
- An additional $450 for each subsequent figure

Permissions
Authors of accepted papers must obtain and provide to the editor on final acceptance all necessary permissions to reproduce in print and electronic form any copyrighted work, including test materials (or portions thereof), photographs, and other graphic images (including those used as stimuli in experiments).
On advice of counsel, APA may decline to publish any image whose copyright status is unknown.
To determine whether you are allowed to reproduce an image, look for the copyright on the work. Here are some examples of copyright statements you might see:

Regular copyright or "all rights reserved" copyright
This kind of copyright is indicated by the word *copyright* or the copyright symbol.
Examples: Copyright 2015 by the American Psychological Association. © 2014 Elizabeth T. Jones, all rights reserved.
Creative commons copyright
Creative commons licenses are indicated by the words *creative commons* or CC. In general, creative commons licenses allow you to reproduce and/or adapt a work (including images) without getting permission from the copyright holder, so long as you give credit to the original author in the form of a copyright statement.

Public domain
Works that are not bound by copyright are considered in the public domain. This means you can reproduce them and/or adapt them however you want, so long as you credit the original author in the form of a copyright statement. Assume a work is under copyright unless you see the words *public domain* on it or the work was produced by the U.S. Government (in which case it is automatically in the public domain). Copyright does expire, but that can take a long time.

No copyright indicated
If no copyright is indicated, treat the work as copyrighted. U.S. copyright law states that a work is copyrighted as soon as it is fixed in tangible form (e.g., you can see it on a computer screen or on paper), even if the work doesn't say copyright or have the copyright symbol, and even if it is not mass produced or professionally published. For example, you automatically own the copyright to papers you write for a class.

Publication Policies
APA policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications.
See also APA Journals® Internet Posting Guidelines.
APA requires authors to reveal any possible conflict of interest in the conduct and reporting of research (e.g., financial interests in a test or procedure, funding by pharmaceutical companies for drug research).
- Download Disclosure of Interests Form (PDF, 38KB)
  In light of changing patterns of scientific knowledge dissemination, APA requires authors to provide information on prior dissemination of the data and narrative interpretations of the data/research appearing in the manuscript (e.g., if some or all were presented at a conference or meeting, posted on a listserv, shared on a website, including academic social networks like ResearchGate, etc.). This information (2–4 sentences) must be provided as part of the Author Note.
  Authors of accepted manuscripts are required to transfer the copyright to APA.
- For manuscripts **not** funded by the Wellcome Trust or the Research Councils UK
  Publication Rights (Copyright Transfer) Form (PDF, 83KB)
- For manuscripts funded by the Wellcome Trust or the Research Councils UK
  Wellcome Trust or Research Councils UK Publication Rights Form (PDF, 34KB)

Ethical Principles

It is a violation of APA Ethical Principles to publish "as original data, data that have been previously published" (Standard 8.13).

In addition, APA Ethical Principles specify that "after research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release" (Standard 8.14).

APA expects authors to adhere to these standards. Specifically, APA expects authors to have their data available throughout the editorial review process and for at least 5 years after the date of publication.

Authors are required to state in writing that they have complied with APA ethical standards in the treatment of their sample, human or animal, or to describe the details of treatment.