COMPASSION-FOCUSED THERAPIES FOR SELF-ESTEEM: A SYSTEMATIC REVIEW AND META-ANALYSIS

Introduction

Self-Esteem

Self-esteem has been defined in a variety of ways over the years (see Mruk (2013) for a comprehensive overview of both global self-esteem and its subtypes). The most common and enduring definition equates global self-esteem with 'global feelings of self-worth': This definition appears to have face validity in describing the phenomenological experience, and consequently will be used for the purposes of this review. Fennell (1997) sees global self-esteem as akin to a self-schema (i.e., a generalised belief about the self that organises self-related information processing), and has identified common characteristics of low self-esteem – such as self-doubt, self-criticism, low sense of entitlement, and perceiving the true self as inadequate or inferior (Fennell, 2009). Fennell (2004) also considered how low self-esteem compares to the negative self-cognition seen in depression, conceptualising this on a continuum of treatment difficulty, with low self-esteem being more difficult to treat, enduring, and associated with greater disability. At the extreme, it is thought to involve "more or less constant self-flagellation…triggered by a wide range of stimuli", with "no counterbalancing positive perspective on the self", "multiple, generalised, longstanding problems, seen as central to the true self", and a "negative perspective on the self seen as reflecting fact" (Fennell, 2004, p. 1060).

Tafarodi and Swann (1995) further characterised global self-esteem in terms of a two-factor model: involving feelings of self-liking but particularly linked to a sense of self-competence within areas of life which are meaningful and/ or of value to an individual. One of the most

commonly used measures of global self-esteem (the Rosenberg Self-Esteem Scale; Rosenberg, 1965) has been shown to reflect this dichotomy (Tafarodi & Swann, 1995).

'Average/ normal' global self-esteem has been associated with psychological wellbeing (Rosenberg, Schooler, Schoenbach, & Rosenberg, 1995) and low global self-esteem has been implicated in a range of psychological difficulties (Silverstone & Salsali, 2003). The direction of the relationship appears unclear, although some research suggests that vulnerability models are supported in relation to anxiety and depression – with low self-esteem identified as a risk factor or contributing factor rather than an outcome (Manna, Falgares, Ingoglia, Como, & De Santis, 2016; Orth & Robins, 2008; Sowislo & Orth, 2013). Kesting and Lincoln (2013) have also proposed a model whereby self-esteem is implicated in the formation and maintenance of persecutory delusions.

Thus, given that low global self-esteem appears to be implicated in the development of a variety of difficulties, tends to be enduring, and has implications for treatment efficacy, it seems pertinent to investigate this (potentially transdiagnostic) construct.

Compassion-focused therapy (CFT)

Compassion-focused therapy is described as a biopsychosocial, evolutionary, and integrative approach based upon Social Mentality Theory (Gilbert, 2010, 2014), and was born out of an understanding of biosocial motivational systems (e.g. related to competition and social rank, caring and cooperation) and the emotion systems which guide them (Gilbert, 2014). Within this context a model of three affect regulation systems (the 'threat', 'drive', and 'soothe' systems) has been proposed, with mental health difficulties seen as arising from over- and/ or under-activity in any one (or a combination) of these systems. For example, depression being conceptualised as a combination of an overactive 'threat' system and underactive 'drive' system. The affect regulation systems are purported to co-regulate each other, however

Gilbert (2014) highlights that some people will have had little opportunity (during early life experiences) to develop the 'soothe' system and associated abilities to regulate the 'threat' and 'drive' systems. Whereas many therapies often work to address 'negative/ threat-based' emotions, CFT works to additionally promote 'positive/ soothe-based' emotions to co-regulate these, particularly using compassion and development of a compassionate mentality.

Gilbert (2014) defines compassion as a "sensitivity to suffering in self and others, with a commitment to try to alleviate and prevent it" (p. 19). A variety of techniques (sometimes referred to as Compassionate Mind Training (CMT)) are associated with the CFT approach, with the aim being to develop compassion (including the skills and attributes associated with this) (Gilbert, 2009). Techniques range, for example, from the use of sensory work (e.g. breathing techniques) to the use of imagery (Gilbert, 2009, 2014).

CFT has primarily been recommended as a means of helping people who experience high levels of shame and self-criticism, and has shown utility for working with people with anxiety, depression, and eating disorders in addition to other non-clinical difficulties (Beaumont & Hollins-Martin, 2015; Goss & Allan, 2014; Kirby, 2017; Kirby, Tellegen, & Steindl, 2017; Leaviss & Uttley, 2015). CFT is described as an approach or focus rather than a new therapy per se and therefore it is important to acknowledge that, although research often makes reference to 'CFT' and may be using compassion-based interventions predicated on similar (compassion-promoting) principles and processes, methodologies may differ.

The Present Review

Although evidence exists for the utility of cognitive behavioural therapy (CBT) tailored for low self-esteem (Kolubinski, Frings, Nikčević, Lawrence, & Spada, 2018), some research (for example that by Stott (2007)) suggests that a cognitive-emotional discrepancy can occur in cognitive therapy whereby, although clients report being able to rationally counter negative

self-beliefs, they find ‘feeling it’ difficult (a finding corroborated by the clinical experience of the first author).

Given that there appear to be themes within the CFT literature that overlap with the concept of low self-esteem (e.g., the importance of the relationship with the self, and an acknowledgment that self-criticism can be associated with perceived weakness or incompetence), it was considered whether CFT or compassion-based interventions might be beneficial for working with low self-esteem. It could perhaps be hypothesised that low self-worth constitutes the cognitive aspect of low self-esteem, with shame and self-criticism (the targets of the CFT approach) being the emotional and behavioural concomitants. There is evidence to suggest that self-criticism can be highly correlated with low self-esteem (Thew, Gregory, Roberts, & Rimes, 2017).

Therefore, the aim of the present review is to identify whether CFT or compassion-based interventions are effective in improving self-esteem.

Method

Search Strategy

A comprehensive search of CINAHL, MEDLINE, and PsycINFO was undertaken, for articles published up to 13th July 2019 (the date of the search). Search terms were primarily synonyms related to the intervention (compassion-focused therapies) and outcome (self-esteem) under study, in the context of PICO criteria. The search was not limited in relation to population or comparator (for the purposes of inclusivity), and thus the comparator could include any comparator, including no comparator. Search terms were entered as truncated free-text terms or identified via the ‘thesaurus’ for each database, in order to ensure breadth of searching, avoidance of bias, and identification of relevant papers.

Handsearching of references and citations of the papers included in the final review (available via PlumX metrics and Scopus) was undertaken, two experts were contacted, and Grey literature (via OpenGrey) and the first ten pages of Google Scholar were also searched, to ensure relevant papers were not missed.

Study Selection

Inclusion and exclusion criteria were used to screen potential papers. Titles and abstracts were initially screened and, if the paper was thought to be eligible (or possibly eligible but unable to be determined from title and abstract alone), the full text was accessed. To be included, papers had to include an intervention identified as being compassion-based and specifically mention "self-esteem" as a variable/ outcome. The latter was with the aim of increasing the clarity and specificity of the phenomenon under study due to historic issues in clearly defining self-esteem, which would also allow for more accurate conclusions to be drawn.

Studies were excluded when they were not in English (unable to be read by the first author), measured appearance-related (rather than global self-) esteem (as per rationale described above), and when self-esteem was included only as a covariate/ predictor (with no pre- and post-assessment to be able to answer the review question). Reviews, theses/ dissertations, study protocols, manuals (e.g., for psychometric tools), and book chapters were also excluded, with the emphasis primarily being on empirical papers relevant to answering the review question. Although review papers were excluded, articles contained within reviews were screened for eligibility.

Data Extraction

Data extraction was undertaken by the first author and included extraction of the following information for each paper: author, year and location, sample details (including sample size,

study or recruitment setting, and primary characteristics such as age and primary sample inclusion criteria or diagnosis), study design, details of intervention and control condition (where applicable), self-esteem measure (including measurement time points), means and standard deviations of pre- and post-intervention scores of self-esteem (with the aim of using these in the meta-analysis), and main outcomes (where relevant to the review question – e.g., the impact of CFT on self-esteem).

Where means and standard deviations for pre- and post-intervention were unavailable, authors were contacted requesting this data. One paper (Murphy, Stosny, & Morrel, 2005) included two studies and samples, only one of which was relevant to the review question (reporting on a sample of participants who attended a compassion-based workshop); therefore data extraction primarily focussed on this particular sample and study.

Quality Appraisal

The Mixed Methods Appraisal Tool (MMAT) Version 2011 (Pluye et al., 2011) was used for the purposes of assessing the methodological quality of included papers given the variety of methodologies across included papers. The aim was to provide a comprehensive appraisal of studies yet one which allowed for clarity and comparison between them (something which may have been difficult if using multiple method-specific tools). The MMAT is designed for the purposes of conducting systematic reviews of mixed methodologies and has demonstrated suitable reliability and validity (Pace et al., 2012).

The tool was adapted for use in the present review in a number of ways. Primarily, since it was felt that some questions were applicable to multiple methodologies (despite being listed only within one methodological section of the MMAT), each study included in the review was assessed against *all* applicable questions in the MMAT. Consequently, the screening

questions were not used to rule out any papers (ensuring inclusivity and thorough quality appraisal of *all* papers, with no papers ruled out on the basis of methodological bias).

Due to adapting the MMAT, this would invalidate the suggested scoring criteria. Therefore, for each included study, average quality scores were calculated by assigning a score of '0', '1' or '2' to responses of "no", "can't tell, and "yes", respectively, summing the scores and then dividing by the number of applicable items/ questions. Screening questions were excluded from these calculations. The aim of the score was to provide additional at-a-glance information regarding the quality of included studies to be interpreted within the wider context of the quality appraisal, rather than being used as a standalone measure.

A sampling frame of 50% was agreed a priori for double-coding by a second reviewer, with any differences of opinion resolved through discussion.

Analysis

The primary planned meta-analysis was of the pooled pre-post effect-size for self-esteem within groups receiving compassion-based intervention. Means and standard deviations for pre- and post-intervention self-esteem data were entered into MetaEssentials 1.4 (Suurmond, van Rhee, & Hak, 2017), along with associated sample sizes. Where means and standard deviations were unavailable, author-reported effect sizes were used (where present), or data were requested from authors. Any papers whereby data remained unavailable were not included in statistical analyses, however these were presented elsewhere in the review to provide relevant context. For one study (Krieger et al., 2019), the effect size was a negative value although mean data and associated description of results suggested an increase (improvement) in self-esteem pre- to post-intervention for the CFT group. The polarity of the effect size was therefore adjusted (to a positive value) for inclusion in analyses. Consistent with recommendations from Balk, Earley, Patel, Trikalinos and Dahabreh (2012) for

calculating within-group effect-sizes, we imputed a value of $r = .59$ for the correlation between pre- and post-intervention self-esteem scores (holding this value constant across studies).

Following calculation of the overall pooled effect size, a forest plot was produced to represent and assess individual and overall effect sizes and confidence intervals, along with a funnel plot to assess small-study or publication bias. Heterogeneity was assessed using the I^2 statistic and a sensitivity analysis conducted to review the effect on heterogeneity of selectively removing included studies such as those with lower quality scores.

A secondary meta-analysis was conducted using Review Manager 5.3 (The Cochrane Collaboration, 2014) to examine the pooled between-groups effect-size for self-esteem in studies with comparator groups (i.e., those comparing a compassion-based intervention group with a control group). Post-intervention means and standard deviations were entered for each study group (intervention and control) along with associated sample sizes to calculate the overall pooled effect size. A forest plot of the data was produced to review individual and overall effect sizes and confidence intervals of included studies, along with a funnel plot to consider small-study or publication bias. Heterogeneity was assessed using the I^2 statistic and a sensitivity analysis conducted to assess the effect on heterogeneity of selectively removing included studies.

For both meta-analyses, a random effects model was utilised due to variation in samples and methodologies. In terms of overall pooled effect size, <0.5 was deemed to be small, 0.5-0.8 medium, and ≥ 0.8 large (in line with standard definitions). For heterogeneity, $\leq 25\%$ was considered to be low, 25-50% moderate, and $\geq 50\%$ high (Higgins, Thompson, Deeks, & Altman, 2003).

Results

Search Strategy

In total, 569 results were identified via CINAHL, MEDLINE and PsycINFO following application of the search strategy, with 338 remaining after removal of duplicates. 256 were ineligible based on screening of titles and abstracts, with a further 72 excluded following full-text review. Results are included in a PRISMA flow diagram (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009) in Figure 1.

Based on the eligibility criteria applied, by screening specifically for "self-esteem" as a phenomenon, papers using associated measures such as those relating to self-compassion and self-criticism were excluded (unless self-esteem was also measured), and interventions which typically were not identified as compassion-based tended to be mindfulness-based stress reduction (MBSR), mindfulness-based cognitive therapy (MBCT) and one intervention of emotion-focused therapy (EFT).

One additional paper was provided by an expert and another identified through handsearching of eligible articles, bringing the total included studies to 10. All reference lists and articles which cited the final 10 included papers were hand-searched. One paper citing one of the included papers initially appeared to meet eligibility criteria, however on reading the full-text (although it mentioned compassion as a component of the intervention) compassion was only identified as a small component and the intervention overall was described as being a CBT intervention. It was therefore excluded on the grounds of not being a compassion-based intervention.

Nine results were produced by searching OpenGrey, however all of these were theses and therefore excluded according to eligibility criteria. The first ten pages of Google Scholar were screened to ensure no key papers were missed. No further papers were identified except one thesis which was again excluded in line with eligibility criteria.

Study Characteristics

Characteristics of included studies are provided in Table 1. From this point onwards studies will be referred to according to their allocated number in Table 1 for clarity and fluency.

Half of the included studies (Studies 2, 3, 5, 7 and 10) were conducted in the UK and the majority (1, 3, 4, 7, 9 and 10) used a quantitative observational design (with two of these (3 and 9) involving retrospective analysis of existing data). Four studies (4-6 and 8) included a comparator group, although only two (6 and 8) involved random allocation to groups, and there was one case study (Study 2). Studies involved a mixture of different samples; with different diagnoses (e.g. anxiety, depression, eating disorders, personality disorders) and different settings (e.g. university setting, outpatient clinic, hospice, maximum security hospital). Sample sizes ranged from 1 (study 2) to 121 (study 6), with a combined sample-size (across studies) of 567; of these, 390 were included in our within-group (pre-post) meta-analyses, and 196 in our between-groups meta-analyses (96 in intervention groups vs 100 in control groups). In order to assess self-esteem as an outcome, the majority used versions of the Rosenberg Self-Esteem Scale (RSE), where this was not used (studies 2, 3 and 5), other measures included the Robson Self-Concept Questionnaire (SCQ), Stirling Eating Disorder Scales (SEDS) self-esteem subscale or the Single-Item Self-Esteem Scale (SISE).

Intervention Characteristics

To be eligible, studies had to include an intervention identified as primarily compassion-focused. Two (in studies 1 and 7) were identified as compassionate mind training (CMT) and five (in studies 2-4, 6 and 10) as compassion-focused therapy (CFT). Of those involving CFT, one (in Study 6) was conducted online and four used CFT in addition to other therapeutic approaches (three with CBT (studies 2, 3 and 10) and one with LKM and MBCT (Study 4)). Other interventions included a workshop format based upon attachment theory

(Study 9) (in contrast to others which referred to Gilbert's work as their influence) and two brief interventions involving a compassion-focused writing task (Study 5) and compassionate image (Study 8). Exposure to, and length of, the interventions varied, with most providing approximately 20-24 hours of intervention (studies 1-4, 7 and 9), two studies (6 and 10) providing approximately 6-7 hours, and in two studies (5 and 8) the intervention lasted a matter of minutes. Homework and/ or practice outside of sessions was encouraged in half of the studies (studies 1-4 and 6), although uptake varied. Intervention length was approximately 12 weeks on average (excluding the two brief intervention studies), and only four (studies 4, 6, 7 and 10) included a follow-up period (ranging from six weeks to six months). Seven studies (1-3 and 7-10) demonstrated a statistically significant increase in self-esteem and/ or clinically-meaningful change pre- to post-intervention (in favour of the compassion-based intervention rather than control group, where applicable), although interestingly in one study (Study 7) this was only true for the RSE and not other measures of self-esteem included in the same study.

Quality Appraisal

Level of agreement between reviewers was assessed and, prior to resolving any differences, overall weighted kappa = .89. Quality appraisal data is presented in Table 2.

Average quality scores ranged between 0.8 and 1.8 with Study 10 scoring the highest and Studies 3 and 5 scoring the lowest. The lowest-scoring studies tended to lack sufficient information to make decisions about quality or possible bias e.g. relating to recruitment strategies (including eligibility criteria) and psychometric properties (such as the reliability and validity) of measures used. They also had higher dropout rates and did not specify a power calculation. Study 5 involved a self-selecting sample and Study 3 involved

retrospective analysis of existing data, hence this appeared to limit the information available regarding sample recruitment and response rate which resulted in a lower quality score.

In terms of common pitfalls when it came to increased possibility of bias or lower quality, this mainly related to questions 1.1, 1.4, 3.1, 3.2, 4.1, 4.2 and 4.4 of the MMAT in terms of sources of data, sample representativeness, sampling and recruitment strategy, acknowledgement of possible researcher influence, use of appropriate measures, and acceptable response rates. This was most commonly due to studies lacking sufficient information regarding recruitment strategy, eligibility criteria, reliability and validity of measures, and power calculation. Studies also rarely demonstrated reflexivity in terms of possible researcher influence or bias. In relation to acceptable response rates, recruitment methods were not necessarily suited to being able to capture this information (e.g. studies did not involve survey methodologies) and therefore reasons for initial non-participation or refusal were not available (although reasons for dropout were at times provided). Overall, the majority of studies did however demonstrate reasonable levels of complete outcome data and low dropout rates.

In addition to the MMAT, it is also worth noting that all self-esteem measures were self-report measures (one of which only included a single item), hence this may be subject to bias, and often studies lacked a control group against which to draw comparisons and conclusions.

Meta-analysis

Primary analysis: Pre-post effect-size

Pooled estimates of pre-post effect-size (examining changes in self-esteem within groups receiving compassion-based interventions) were computed based on data from eight of ten studies – two small-N studies (2, 5) could not be included, as they provided insufficient data to calculate effect sizes.

Hedges' *g* was used to calculate overall effect size. The results demonstrated a medium, significant overall effect size, with self-esteem increasing following compassion-based intervention (g = 0.56, 95% CI [0.19-0.93], z = 3.54, $p < .001$). Statistical heterogeneity was significantly high (I^2 = 85.70%, $p < .001$) indicating that effect estimates were variable and likely influenced by overt clinical and methodological differences. A forest plot was produced (see Figure 2) which suggested that five of eight studies showed a significant pre-post increase in self-esteem. A funnel plot suggested there were no marked issues relating to small-study or publication bias (see Figure 3), although study 6 was a notable outlier (with CIs that did not overlap with the pooled CIs) and its influence was consequently examined further in a secondary sensitivity analysis (reported below).

Sensitivity analysis.

Sensitivity analysis was conducted to review the impact on heterogeneity and effect size of excluding the two lowest quality studies (Studies 3 and 9). By removing Studies 3 and 9, the overall effect-size increased slightly and remained significant (g = 0.61, 95% CI [0.05-1.17], z = 2.82, p = .005); although CIs widened (reflecting a loss of power) and heterogeneity also increased slightly (I^2 = 88.05%, $p < .001$).

In a secondary (post-hoc) sensitivity analysis, excluding the study with the briefest intervention (8) led to a slight increase in overall effect-size (g = 0.62, 95% CI [0.20-1.04], z = 3.64, $p < .001$); heterogeneity remained high (I^2 = 85.74%, $p < .001$). A further secondary sensitivity analysis, excluding the above-identified outlier study (6), led to a decreased overall effect-size (g = 0.45, 95% CI [0.14-0.76], z = 3.59, $p < .001$) with slightly reduced heterogeneity (I^2 = 72.99%, p = .001).

Secondary analysis: Between-groups effect-size.

A planned secondary analysis examined whether effect-sizes favoured compassion-based interventions when comparing against control conditions. Study 4 was excluded from these analyses since groups were dependent (a baseline wait list control) rather than independent. Studies 5, 6, and 8 were included in the analysis, which resulted in a non-significant overall medium effect size (g = 0.58, 95% CI [-0.08-1.24], z = 1.73, p = 0.08). Heterogeneity was again significantly high in this analysis (I^2 = 69%, p = 0.04). The associated forest plot (see Figure 4) showed that Study 6 demonstrated an effect size significantly in favour of the compassion-based intervention, whereas the other two studies were non-significant, likely due to the latter being very brief interventions. The confidence interval of Study 5 was fairly large (-1.34-2.64) suggesting the study lacks reliability. By removing either of the other two studies, heterogeneity fell to zero but the large overlapping confidence interval of Study 5 appeared to account for this. The funnel plot did not appear to provide any meaningful data given that this included only three data points.

Discussion

Within-group analysis of available data indicated a significant, medium combined pre-post effect-size, which was robust to applied sensitivity analyses, suggesting that compassion-based interventions may be effective in improving self-esteem. A similar (medium) pooled effect-size was observed for studies comparing compassion-based intervention against a control condition, although this effect was not statistically significant (reflecting the small numbers in [lower precision afforded by] these controlled studies). Whilst not included in pooled analyses, the results of Ashworth, Gracey, and Gilbert's (2011) single case study were consistent with meta-analytic findings – with CFT promoting a clinically meaningful change in self-esteem for a client with an acquired brain injury. Large heterogeneity amongst studies does, however, make it difficult to draw further conclusions. In reviewing forest plots for (both pre-post and between-groups) meta-analyses, four (of ten) individual studies (Graser,

Höfling, Weßlau, Mendes, & Stangier, 2016; Imrie & Troop, 2012; Laithwaite et al., 2009; Lincoln, Hohenhaus, & Hartmann, 2013) showed non-significant results. Notably, two of these studies applied brief interventions – one of which (Imrie & Troop, 2012) demonstrated a particularly large confidence interval suggesting the study may lack reliability. Indeed, this included a very small sample, applied a brief intervention (something acknowledged by the authors due to practical difficulties), and used a single-item self-report measure of self-esteem – scoring low in terms of quality criteria. In terms of intervention length, it seems particularly important in compassion-based interventions to allow sufficient time for intervention and follow-up. It has often been reported that clients with mental health difficulties may initially struggle to foster self-compassion and even find this uncomfortable or distressing (associated with early life experiences lacking in sources of 'soothe' and compassion and perhaps characterised by trauma (Gilbert, 2009)). Hence, sufficient time is required to understand and formulate these difficulties in order to effectively nurture self-compassion. This may go some way in explaining why Graser et al.'s (2016) study did not demonstrate significant intervention effects on self-esteem, as some participants anecdotally reported that compassion triggered negative emotions.

The measurement of self-esteem in itself has its challenges since the majority of tools tend to rely on self-report (exceptions sometimes being measures of implicit self-esteem used in studies, such as the implicit association test (Kim & Moore, 2019)). Interestingly in Laithwaite et al.'s (2009) study, there was a significant improvement in self-esteem measured using the RSE but not for the SCQ or SIP-AD-AE, perhaps related to differing psychometric properties. The RSE scale does however appear to remain the most widely used.

Recommendations for Future Research

Given the scarcity of available literature, the present review had to be inclusive, however this resulted in clinical and methodological heterogeneity. Ideally, future studies would include randomised controlled trials with comparable control groups, with sufficient intervention length (a total intervention duration ≥ 20 hours, with weekly contacts, seemed appropriate within the current review), and sufficient follow-up. However, given that CFT is identified as an 'approach' rather than 'intervention' per se (which can be incorporated into various therapeutic modalities), it may not be possible to do this within naturalistic settings to determine the contribution of individual therapeutic components. Perhaps an alternative is that more studies in general are needed, which can then be reviewed on a group rather than individual level – e.g. by performing sub-group analyses on similar 'clusters' of methodologies to tease out effects.

In a recent meta-analytic review, CBT approaches for improving self-esteem (based upon Fennell's work) demonstrated a large pooled effect size (Kolubinski et al., 2018). However, it is worthwhile noting that included studies and interventions had a primary focus on self-esteem (unlike those in the present review whereby self-esteem was not always a primary target), and the review excluded some participants with more severe mental health difficulties. In the future it would be useful to directly compare the efficacy of different treatment approaches in improving self-esteem. Compassion-based approaches, for example, may offer additional therapeutic benefit where the effect of first-line CBT has been limited (similar to Studies 2 and 3 included in the present review), perhaps in instances of cognitive-emotional discrepancy (Stott, 2007). Neff (2003) has drawn distinctions between self-esteem and self-compassion, proposing that the latter is better able to cope with failure or setbacks and is not specifically linked to self-evaluation, with Neff and Vonk (2009) having suggested that self-compassion offers a healthier alternative to self-esteem. Consequently, whilst CBT may directly target and improve self-esteem, CFT may offer an indirect approach for

improving self-esteem (by fostering compassion), with the potential for long-term benefits as a buffer against the psychological and emotional impact of future setbacks and adversities.

Conclusion

The present review demonstrated a significant, medium combined effect size for the within-group (pre-to-post) effect of compassion-based interventions on self-esteem. This suggests that CFT and related compassion-based approaches may be effective in improving self-esteem – and contributes to existing literature regarding the transdiagnostic efficacy of compassion-based intervention. Heterogeneity between studies was however high, making it difficult to draw further conclusions. It seems pertinent to consider transdiagnostic concepts (such as low self-esteem and CFT or compassion-based intervention), at the level of core difficulties, in supporting people with their mental health – particularly perhaps when challenges are encountered in therapy (e.g., treatment resistance or cognitive-emotional discrepancy). This may foster the therapeutic alliance, support longer-term therapeutic benefit, but also reduce the need for the development of diagnosis-specific treatment models where appropriate.

References

Andersen, B., & Rasmussen, P. H. (2017). Transdiagnostic group therapy for people with self- critic and low self esteem, based on compassion focused therapy principles. *Journal of Compassionate Health Care, 4*(14).

Ashworth, F., Gracey, F., & Gilbert, P. (2011). Compassion focused therapy after traumatic brain injury: Theoretical foundations and a case illustration. *Brain Impairment, 12*(2), 128-139. doi:10.1375/brim.12.2.128

Balk, E. M., Earley, A., Patel, K., Trikalinos, T. A., Dahabreh, I. J. (2012). *Empirical Assessment of Within-arm Correlation Imputation in Trials of Continuous Outcomes* (Methods Research Report No. AHRQ Publication No. 12 (13)-EHC141-EF). Agency for Healthcare Research and Quality: Rockville, MD. Retrieved from www.effectivehealthcare.ahrq.gov/reports/final.cfm

Beaumont, E., & Hollins-Martin, C. J. (2015). A narrative review exploring the effectiveness of Compassion-Focused Therapy. *Counselling Psychology Review, 30*(1), 21-32.

Fennell, M. J. V. (1997). Low Self-Esteem: A Cognitive Perspective. *Behavioural and Cognitive Psychotherapy, 25*, 1-25.

Fennell, M. J. V. (2004). Depression, low self-esteem and mindfulness. *Behaviour Research and Therapy, 42*, 1053–1067.

Fennell, M. J. V. (2009). *Overcoming low self-esteem: A self-help guide using Cognitive Behavioral Techniques*. London: Robinson.

Gale, C., Gilbert, P., Read, N., & Goss, K. (2014). An evaluation of the impact of introducing compassion focused therapy to a standard treatment programme for people with eating disorders. *Clinical Psychology & Psychotherapy, 21*(1), 1-12. doi:10.1002/cpp.1806

Gilbert, P. (2009). Introducing compassion-focused therapy. *Advances in Psychiatric Treatment, 15*, 199-208. doi:10.1192/apt.bp.107.005264

Gilbert, P. (2010). *Compassion focused therapy: Distinctive features*. New York, NY: Routledge/Taylor & Francis Group.

Gilbert, P. (2014). The origins and nature of compassion focused therapy. *British Journal of Clinical Psychology, 53*(1), 6-41. doi:10.1111/bjc.12043

Goss, K., & Allan, S. (2014). The development and application of compassion-focused therapy for eating disorders (CFT-E). *British Journal of Clinical Psychology, 53*(1), 62-77. doi:10.1111/bjc.12039

Graser, J., Höfling, V., Weßlau, C., Mendes, A., & Stangier, U. (2016). Effects of a 12-week mindfulness, compassion, and loving kindness program on chronic depression: A pilot within-subjects wait-list controlled trial. *Journal of Cognitive Psychotherapy, 30*(1), 35-49. doi:10.1891/0889-8391.30.1.35

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ (Clinical Research Ed.), 327*(7414), 557-560.

Imrie, S., & Troop, N. A. (2012). A pilot study on the effects and feasibility of compassion-focused expressive writing in Day Hospice patients. *Palliative & Supportive Care, 10*(2), 115-122. doi:10.1017/S1478951512000181

Jones, C. E. (2011). *An Exploration of the Construct Validity of Self-Compassion.* (DClinPsych in Clinical and Community Psychology). University of Exeter,

Kesting, M.-L., & Lincoln, T. M. (2013). The relevance of self-esteem and self-schemas to persecutory delusions: A systematic review. *Comprehensive Psychiatry, 54*(7), 766-789. doi:10.1016/j.comppsych.2013.03.002

Kim, H. S., & Moore, M. T. (2019). Symptoms of depression and the discrepancy between implicit and explicit self-esteem. *Journal of Behavior Therapy and Experimental Psychiatry, 63*, 1-5. doi:10.1016/j.jbtep.2018.12.001

Kirby, J. N. (2017). Compassion interventions: The programmes, the evidence, and implications for research and practice. *Psychology and Psychotherapy: Theory, Research and Practice, 90*(3), 432-455. doi:10.1111/papt.12104

Kirby, J. N., Tellegen, C. L., & Steindl, S. R. (2017). A meta-analysis of compassion-based interventions: Current state of knowledge and future directions. *Behavior Therapy, 48*(6), 778-792. doi:10.1016/j.beth.2017.06.003

Kolubinski, D. C., Frings, D., Nikčević, A. V., Lawrence, J. A., & Spada, M. M. (2018). A systematic review and meta-analysis of CBT interventions based on the Fennell model of low self-esteem. *Psychiatry Research, 267*, 296-305. doi:10.1016/j.psychres.2018.06.025

Krieger, T., Reber, F., von Glutz, B., Urech, A., Moser, C. T., Schulz, A., & Berger, T. (2019). An Internet-Based Compassion-Focused Intervention for Increased Self-Criticism: A Randomized Controlled Trial. *Behavior Therapy, 50*(2), 430-445. doi:10.1016/j.beth.2018.08.003

Laithwaite, H., O'Hanlon, M., Collins, P., Doyle, P., Abraham, L., Porter, S., & Gumley, A. (2009). Recovery After Psychosis (RAP): a compassion focused programme for individuals residing in high security settings. *Behavioural And Cognitive Psychotherapy, 37*(5), 511-526. doi:10.1017/S1352465809990233

Leaviss, J., & Uttley, L. (2015). Psychotherapeutic benefits of compassion-focused therapy: An early systematic review. *Psychological Medicine, 45*(5), 927-945. doi:10.1017/S0033291714002141

Lincoln, T. M., Hohenhaus, F., & Hartmann, M. (2013). Can paranoid thoughts be reduced by targeting negative emotions and self-esteem? An experimental investigation of a brief compassion-focused intervention. *Cognitive Therapy and Research, 37*(2), 390-402. doi:10.1007/s10608-012-9470-7

Manna, G., Falgares, G., Ingoglia, S., Como, M. R., & De Santis, S. (2016). The Relationship between Self-Esteem, Depression and Anxiety: Comparing Vulnerability and Scar Model in the Italian Context *Mediterranean Journal of Clinical Psychology, 4*(3), 1-17.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med, 6*. doi:10.1371/journal.pmed1000097

Mruk, C. J. (2013). *Self-Esteem and Positive Psychology: Research, Theory, and Practice* (4th ed.). New York, NY: Springer Publishing Company.

Murphy, C. M., Stosny, S., & Morrel, T. M. (2005). Change in self-esteem and physical aggression during treatment for partner violent men. *Journal of Family Violence, 20*(4), 201-210. doi:10.1007/s10896-005-5983-0

Neff, K. D. (2003). Self-Compassion: An Alternative Conceptualization of a Healthy Attitude Toward Oneself. *Self and Identity, 2*(2), 85-101. doi:10.1080/15298860309032

Neff, K. D., & Vonk, R. (2009). Self-compassion versus global self-esteem: Two different ways of relating to oneself. *Journal of Personality, 77*(1), 23-50. doi:10.1111/j.1467-6494.2008.00537.x

Orth, U., & Robins, R. W. (2008). Low Self-Esteem Prospectively Predicts Depression in Adolescence and Young Adulthood. *Journal of Personality and Social Psychology, 95*, 695-708.

Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J., & Seller, R. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies, 49*(1), 47-53. doi:10.1016/j.ijnurstu.2011.07.002

Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'Cathain, A., Griffiths, F., . . . Rousseau, M. C. (2011). Proposal: A mixed methods appraisal tool for systematic mixed studies reviews. *Department of Family Medicine, McGill University, Montreal, Canada.* Retrieved on 8th of August 2019 from http://mixedmethodsappraisaltoolpublic.pbworks.com.

Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and ageing, 17*(3), 423-434.

Rose, A., McIntyre, R., & Rimes, K. A. (2018). Compassion-focused intervention for highly self-critical individuals: Pilot study. *Behavioural and Cognitive Psychotherapy, 46*(5), 583-600. doi:10.1017/S135246581800036X

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review, 60*(1), 141-156. doi:10.2307/2096350

Silverstone, P. H., & Salsali, M. (2003). Low self-esteem and psychiatric patients: Part I – The relationship between low self-esteem and psychiatric diagnosis. *Annals of General Hospital Psychiatry, 2*(1), 1-9.

Sowislo, J. F., & Orth, U. (2013). Does Low Self-Esteem Predict Depression and Anxiety? A Meta-Analysis of Longitudinal Studies *Psychological Bulletin, 139*(1), 213-240.

Stott, R. (2007). When head and heart do not agree: A theoretical and clinical analysis of Rational-Emotional Dissociation (RED) in cognitive therapy. *Journal of Cognitive Psychotherapy, 21*(1), 37-50. doi:10.1891/088983907780493313

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison and validation of Meta-Essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods, 8*, 537-553. doi:https://doi.org/10.1002/jrsm.1260

Tafarodi, R. W., & Swann, W. B. (1995). Self-Linking and Self-Competence as Dimensions of Global Self-Esteem: Initial Validation of a Measure. *Journal of Personality Assessment, 65*(2), 322-342.

The Cochrane Collaboration. (2014). Review Manager (RevMan) [Computer program] (Version 5.3). Copenhagen: The Nordic Cochrane Centre.

Thew, G. R., Gregory, J. D., Roberts, K., & Rimes, K. A. (2017). The phenomenology of self-critical thinking in people with depression, eating disorders, and in healthy individuals. *Psychology and Psychotherapy: Theory, Research and Practice, 90*(4), 751-769. doi:10.1111/papt.12137

Tables

Table 1

Characteristics of included studies

No.	Study, location	Sample (n, setting, characteristics)	Design	Intervention	Control	Measure and data points	Mean (standard deviation)	Main outcomes
1	Andersen and Rasmussen (2017), Denmark	n=102 (10 groups of 10-12 participants) Setting: Private psychiatric practice Characteristics: Age 20-69, 84% female, ICD10 diagnoses: 36% affective disorder, 35% anxiety disorder, 29% personality disorder. Some met criteria for both axis one and axis two diagnoses.	Observational	CMT programme 10 x two-hour weekly sessions Follow-up session after one or two months Included daily homework	N/A	Measure: RSE Timepoints: Pre and post treatment (1st and last session) 75 participants with complete data	(n=75) Pre: 10.61 (5.15) Post: 15.15 (6.54) T -6.70** (p<.001) Cohen's d 0.75 (moderate effect)	Significant increase in self-esteem (moderate to large effect size), more so for those with diagnosis of personality disorder. Therapy effective regardless of diagnosis. People with low self-esteem benefited most from intervention.
2	Ashworth, Gracey and Gilbert (2011), UK	n=1 Setting: holistic neuro-psychological rehabilitation program Characteristics: age 23, diagnosis of traumatic brain injury, alongside self-reported difficulties with anxiety, depression, low self-esteem, anger, and disordered eating.	Case study	24-week neurorehabilitation programme: Initially CBT-based (for first six weeks) before being changed to CFT. (Therapy also adapted in context of ABI). 24 weekly x 50-minute sessions Practice encouraged outside of sessions.	N/A	Measure: SCQ Timepoints: Pre and post treatment (weeks 1 and 24)	Pre: 97 (raw score) Post: 124 (raw score) RCI 8.96 (reliable change)	Clinically meaningful increase in self-esteem based on Reliable Change Index. Belief "I am worthless" fell from 100% to 10%. 'Perfect nurturer' (identified by participant as key catalyst to change), reported to be helpful in terms of returning to work. Felt much 'lighter' in self, repetitive training seen as key along with having meaningful 'perfect nurturer' they could 'feel'.

Table 1 cont.

No.	Study, location	Sample (n, setting, characteristics)	Design	Intervention	Control	Measure and data points	Mean (standard deviation)	Main outcomes
3	Gale, Gilbert, Read and Goss (2014), UK	n=99 Setting: Coventry Eating Disorders Service Characteristics: age 17-62 (mean 28.01), 95 females and four males, 54.5% primary diagnosis of EDNOS, 19.2% anorexia nervosa, 26.3% bulimia nervosa.	Observational retrospective	Group-based two-step treatment programme: Psychoeducation programme (four x two-hour weekly sessions), followed by Recovery Programme based on CBT with subsequent addition of CFT (20 x 2-2.5-hour sessions, twice weekly in first four weeks, weekly thereafter up to 16 weeks). Homework (two hours per week)	N/A	Measure: SEDS self-esteem subscale Timepoints: Initial assessment (T1), pre-psychoeducation programme (T2), post-psychoeducation programme/ pre-recovery programme (T3), after session 8 of recovery programme (T4), end of recovery programme (T5)	(n=62) T1: 23.52 (9.34) T5: 18.27 (11.66)	Significant main effect of time (T1 to T5) for SEDS self-esteem subscale.
4	Graser, Höfling, Weßlau, Mendes and Stangier (2016), Germany	n=11 Setting: Outpatient clinic Characteristics: people with chronic depression (10 diagnosed with double depression, one dysthymia), four female, seven male, age 34-68 (mean 46.46), mean duration of depression 11.55 years.	Observational	Mindfulness (MBCT), compassion and loving-kindness group therapy program (12 x 100-minute weekly sessions) 30-minutes practice x six days per week encouraged	Wait list control (three-month wait period between baseline and treatment)	Measure: German version of RSE Timepoints: baseline, end of 12-week treatment, three-month follow-up.	(n=10) Pre: 24.90 (5.51) Post: 25.00 (6.10)	No significant change in self-esteem over time. Some participants anecdotally reported concentration on compassion for suffering of selves triggered negative affects.

Table 1 cont.

No.	Study, location	Sample (n, setting, characteristics)	Design	Intervention	Control	Measure and data points	Mean (standard deviation)	Main outcomes
5	Imrie and Troop (2012), UK	n=13 (tested in groups of two to seven participants) Setting: day hospice for people with life-limiting illnesses Characteristics: age 38-86 (mean 67.5), eight female, five male, people with life-limiting illnesses	Non-RCT	Writing about stressful experiences WITH self-compassionate expressive writing instruction (introduced ten minutes into 20-minute writing task) Writing task completed during week one and week two, with post-intervention assessment at week three.	Writing about stressful experiences WITHOUT self-compassionate expressive writing instruction Writing task completed during week one and week two, with post-intervention assessment at week three.	Measure: SISE Timepoints: Pre and post (weeks one and three) Five participants with complete data	Raw scores (pre, post) for experimental group: Participant 1: 4, N/A Participant 2: 2, 3 Participant 3: 2, 4 Raw scores (pre, post) for control group: Participant 1: 4, 4 Participant 2: 3, 2 Participant 3: 3, 1	Based upon raw scores, self-esteem in the intervention group increased as compared to the control group. In the compassionate group, two participants' self-esteem increased and one did not complete follow-up. In the stress-only group, two participants' self-esteem decreased and the third was unchanged.
6	Krieger et al. (2019), Switzerland	n=121 (59 in intervention group, 62 in control group) Setting: online Characteristics: Self-critical people, mean age 37.4 (control group) and 37.98 (intervention group), majority female, university-educated, 13.2% current major depressive episode, 34.7% current anxiety disorder.	RCT	Internet-based compassion-focused intervention PLUS Care as usual Included seven modules (50-60 minutes each, one per week encouraged), with up to eight weeks available for completion. Encouraged to complete exercises/ diaries as often as possible. Guidance/ assistance from Psychologist available on request.	Care as usual NB: Offered opportunity for intervention after post-assessment.	Measure: German version of RSE Timepoints: Baseline (pre-treatment), eight weeks (post-treatment), six-months (follow-up for intervention group only). 107 participants completed post-treatment.	Intervention: Pre 1.21 (0.44), post 1.89 (0.55). Control: Pre 1.27 (0.52), post 1.36 (0.56).	Between-group effect size of self-esteem was in favour of intervention group. Within-group effect size was in expected direction.

Table 1 cont.

No.	Study, location	Sample (n, setting, characteristics)	Design	Intervention	Control	Measure and data points	Mean (standard deviation)	Main outcomes
7	Laithwaite et al. (2009), UK	n=19 (Three groups) Setting: maximum security NHS hospital Characteristics: All male, mean age 36.9, diagnoses: schizophrenia, paranoid schizophrenia, or bipolar affective disorder with a history of auditory hallucinations when elated, eight participants had a co-morbid personality disorder (primarily anti-social).	Observational	Recovery group based upon CMT (Recovery After Psychosis (RAP) programme) Two sessions per week x 10 weeks (20 sessions) Follow-up after six weeks Between-session individual support available	N/A	Measure: RSE, SCQ and SIP-AD-SE Timepoints: At the start, middle (week five), end, and at six-week follow-up. 18 participants with complete data.	(n=18) Only median and IQR available	Significant improvement in self-esteem (from baseline to six-week follow-up) based upon results of RSE. No significant change based upon results of SCQ or SIP-AD-SE.
8	Lincoln, Hohenhaus and Hartmann (2013), Germany	n=71 (35 subsequently included in each group) Setting: researcher office Characteristics: healthy participants with subclinical levels of psychosis, all students, age 18-50 (mean 23.2), 69% female, six with previous diagnosis of mental disorder.	RCT	Brief compassion-focused intervention (use of a compassion-focused image following negative emotion induction)	Neutral control condition (use of a neutral image following negative emotion induction)	Measure: state-adapted RSE Timepoints: Baseline, Pre-intervention, mid-intervention (after negative emotion induction), and post-intervention 70 participants with complete data.	Intervention: Pre 2.2 (0.6), post 2.3 (0.6). Control: Pre 2.3 (0.6), post 2.2 (0.6)	Condition had significant effect on self-esteem. Participants in compassion condition reported higher self-esteem than those in control condition.

Table 1 cont.

No.	Study, location	Sample (n, setting, characteristics)	Design	Intervention	Control	Measure and data points	Mean (standard deviation)	Main outcomes
9	Murphy, Stosny and Morrel (2005), US	n=107 Setting: domestic violence treatment program Characteristics: male domestic abuse perpetrators, age 23-70 (mean 40.1), mean education 13.4 years, 91% court-referred for treatment, 72% African-American	Observational retrospective	Compassion workshop based on attachment theory formulation 12 x two-hour weekly sessions	N/A	Measure: RSE (though used with a five-point likert scale, maximum 50 points) Timepoints: Pre-treatment and post-treatment (during last session)	Pre: 41.4 (6.4) Post: 43.5 (4.7)	Significant increase in self-esteem pre- to post-treatment (small effect size).
10	Rose, McIntyre and Rimes (2018), UK	n=23 Setting: university population, questionnaire data completed online Characteristics: university students (majority post-graduate) with high self-criticism causing clinically-significant functional impairment (according to Work and Social Adjustment Scale), mean age 25.3, 82.6% female, 73.9% Caucasian, 56.5% past diagnosis of depression	Observational	Compassion-focused intervention, including general CBT principles 6 x one-hour (approximately weekly) sessions Written booklets provided	N/A	Measure: RSE Timepoints: Sessions one, three (mid-treatment), six (post-treatment), and at two-month follow-up	(n=21) Pre: 13.22 (3.95) Post: 17.57 (3.79) Cohen's d 1.10 (post-pre)	Statistically significant improvement in self-esteem between pre- and post-intervention. Small effect size over baseline period (non-significant), large effect size at post-treatment and follow-up.

ABI, acquired brain injury; CBT, Cognitive-behavioural therapy; CFT, Compassion-focused therapy, CMT, Compassionate Mind Training; EDNOS, eating disorder not otherwise specified; ICD-10, International Statistical Classification of Diseases and Related Health Problems 10th Revision; MBCT, mindfulness-based cognitive therapy; NHS, National Health Service; RCI, reliable change index; RCT, randomised controlled trial; RSE, Rosenberg Self-Esteem Scale; SCQ, Robson Self-Concept Questionnaire; SEDS, Stirling Eating Disorder Scales; SIP-AD-SE, Self-Image Profile for Adults self-esteem subscale; SISE, Single-Item Self-Esteem Scale.

Table 2

Quality appraisal of included studies (based on MMAT)

	No. 1 Andersen and Rasmussen (2017), Denmark	No. 2 Ashworth, Gracey and Gilbert (2011), UK	No. 3 Gale, Gilbert, Read and Goss (2014), UK	No. 4 Graser, Höfling, Weßlau, Mendes and Stangier (2016), Germany	No. 5 Imrie and Troop (2012), UK	No. 6 Krieger et al. (2019), Switzerland	No. 7 Laithwaite et al. (2009), UK	No. 8 Lincoln, Hohenhaus and Hartmann (2013), Germany	No. 9 Murphy, Stosny and Morrel (2005), US	No. 10 Rose, McIntyre and Rimes (2018), UK
Screening question: Are there clear qualitative and quantitative research questions (or objectives), or a clear mixed methods question (or objective)?	N/A	N/A	N/A	N/A	N/A	Yes	N/A	N/A	N/A	Yes
Screening question: Do the collected data allow address the research question (objective)?	Yes	Yes	Yes	Can't tell	No	Yes	Can't tell	Yes	Yes	Yes
1.1. Are the sources of *data* (archives, documents, informants, observations) relevant to address the research question (objective)?	Yes	Yes	Can't tell	Yes	Can't tell	Can't tell	Can't tell	Can't tell	Can't tell	Yes

Table 2 cont.

	No. 1 Andersen and Rasmussen (2017), Denmark	No. 2 Ashworth, Gracey and Gilbert (2011), UK	No. 3 Gale, Gilbert, Read and Goss (2014), UK	No. 4 Graser, Höfling, Weßlau, Mendes and Stangier (2016), Germany	No. 5 Imrie and Troop (2012), UK	No. 6 Krieger et al. (2019), Switzerland	No. 7 Laithwaite et al. (2009), UK	No. 8 Lincoln, Hohenhaus and Hartmann (2013), Germany	No. 9 Murphy, Stosny and Morrel (2005), US	No. 10 Rose, McIntyre and Rimes (2018), UK
1.2 Is the process for analysing *data* relevant to address the research question (objective)?	Yes	Yes	Yes	Can't tell	Can't tell	Can’t tell	No	Yes	No	Yes
1.3 Is appropriate consideration given to how findings relate to the context, e.g., the setting, in which the data were collected?	No	Yes	Can't tell	Can't tell	Can't tell	Yes	Can't tell	Yes	No	Yes
1.4 Is appropriate consideration given to how findings relate to researchers' influence, e.g., through their interactions with participants?	Can't tell	No	No	Yes	Can't tell	No	Can't tell	No	No	Yes
2.1 (If RCT) Is there a clear description of the randomisation (or an appropriate sequence generation)?	N/A	N/A	N/A	N/A	N/A	Yes	N/A	Can't tell	N/A	N/A

Table 2 cont.

	No. 1 Andersen and Rasmussen (2017), Denmark	No. 2 Ashworth, Gracey and Gilbert (2011), UK	No. 3 Gale, Gilbert, Read and Goss (2014), UK	No. 4 Graser, Höfling, Weßlau, Mendes and Stangier (2016), Germany	No. 5 Imrie and Troop (2012), UK	No. 6 Krieger et al. (2019), Switzerland	No. 7 Laithwaite et al. (2009), UK	No. 8 Lincoln, Hohenhaus and Hartmann (2013), Germany	No. 9 Murphy, Stosny and Morrel (2005), US	No. 10 Rose, McIntyre and Rimes (2018), UK
2.2. (If RCT) Is there a clear description of the allocation concealment (or blinding when applicable)?	N/A	N/A	N/A	N/A	N/A	Can't tell	N/A	No	N/A	N/A
2.3. Are there complete outcome data (80% or above)?	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes
2.4. Is there low withdrawal/ drop-out (below 20%)?	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Can't tell	Yes
3.1. Are participants (organisations) recruited in a way that minimises selection bias?	Can't tell	N/A	Can't tell	Yes	Can't tell	Can't tell	Can't tell	Can't tell	Can't tell	Yes

Table 2 cont.

	No. 1 Andersen and Rasmussen (2017), Denmark	No. 2 Ashworth, Gracey and Gilbert (2011), UK	No. 3 Gale, Gilbert, Read and Goss (2014), UK	No. 4 Graser, Höfling, Weßlau, Mendes and Stangier (2016), Germany	No. 5 Imrie and Troop (2012), UK	No. 6 Krieger et al. (2019), Switzerland	No. 7 Laithwaite et al. (2009), UK	No. 8 Lincoln, Hohenhaus and Hartmann (2013), Germany	No. 9 Murphy, Stosny and Morrel (2005), US	No. 10 Rose, McIntyre and Rimes (2018), UK
3.2. Are measurements of review-focal outcome (self-esteem) appropriate (clear origin, or validity known), or standard instrument; and absence of contamination between groups (when appropriate) regarding the exposure/ intervention and outcomes?	Yes	Can't tell	Can't tell	Yes	Can't tell	Yes	Can't tell	Can't tell	Can't tell	Yes
3.3 (If group-comparative) In the groups being compared (exposed vs. non-exposed; with intervention vs. without; cases vs. controls), are the participants comparable, or do researchers take into account (control for) the difference between these groups?	N/A	N/A	N/A	N/A	Can't tell	Can't tell	N/A	Yes	N/A	N/A

Table 2 cont.

	No. 1 Andersen and Rasmussen (2017), Denmark	No. 2 Ashworth, Gracey and Gilbert (2011), UK	No. 3 Gale, Gilbert, Read and Goss (2014), UK	No. 4 Graser, Höfling, Weßlau, Mendes and Stangier (2016), Germany	No. 5 Imrie and Troop (2012), UK	No. 6 Krieger et al. (2019), Switzerland	No. 7 Laithwaite et al. (2009), UK	No. 8 Lincoln, Hohenhaus and Hartmann (2013), Germany	No. 9 Murphy, Stosny and Morrel (2005), US	No. 10 Rose, McIntyre and Rimes (2018), UK
3.4. Are there complete outcome data (80% or above), and, when applicable, an acceptable response rate (60% or above), or an acceptable follow-up rate for cohort studies (depending on the duration of follow-up)?	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes
4.1. Is the sampling strategy relevant to address the *research question*?	Can't tell	Yes	Can't tell	Can't tell	Can't tell	Yes	Can't tell	Can't tell	Can't tell	Yes
4.2. Is the sample representative of the population under study?	Yes	Yes	Can't tell	Yes	Can't tell	Can't tell	Can't tell	Can't tell	Can't tell	Yes

Table 2 cont.

	No. 1 Andersen and Rasmussen (2017), Denmark	No. 2 Ashworth, Gracey and Gilbert (2011), UK	No. 3 Gale, Gilbert, Read and Goss (2014), UK	No. 4 Graser, Höfling, Weßlau, Mendes and Stangier (2016), Germany	No. 5 Imrie and Troop (2012), UK	No. 6 Krieger et al. (2019), Switzerland	No. 7 Laithwaite et al. (2009), UK	No. 8 Lincoln, Hohenhaus and Hartmann (2013), Germany	No. 9 Murphy, Stosny and Morrel (2005), US	No. 10 Rose, McIntyre and Rimes (2018), UK
4.3. Are measurements of review-focal outcome (self-esteem) appropriate (clear origin, or validity known, or standard instrument)?	Yes	Can't tell	Can't tell	Yes	Can't tell	Yes	Can't tell	Can't tell	Can't tell	Yes
4.4. Is there an acceptable response rate (60% or above)?	Can't tell	Yes	Yes	Can't tell	Can’t tell	Can't tell	Can't tell	Can't tell	Can't tell	Can't tell
5.1. (If mixed methods) Is the mixed methods research design relevant to address the qualitative and quantitative research questions (or objectives), or the qualitative and quantitative aspects of the mixed methods question (or objective)?	N/A	N/A	N/A	N/A	N/A	Yes	N/A	N/A	N/A	Yes

Table 2 cont.

	No. 1 Andersen and Rasmussen (2017), Denmark	No. 2 Ashworth, Gracey and Gilbert (2011), UK	No. 3 Gale, Gilbert, Read and Goss (2014), UK	No. 4 Graser, Höfling, Weßlau, Mendes and Stangier (2016), Germany	No. 5 Imrie and Troop (2012), UK	No. 6 Krieger et al. (2019), Switzerland	No. 7 Laithwaite et al. (2009), UK	No. 8 Lincoln, Hohenhaus and Hartmann (2013), Germany	No. 9 Murphy, Stosny and Morrel (2005), US	No. 10 Rose, McIntyre and Rimes (2018), UK
5.2. (If mixed methods) Is the integration of qualitative and quantitative data (or results) relevant to address the research question (objective)?	N/A	N/A	N/A	N/A	N/A	Yes	N/A	N/A	N/A	Yes
5.3. (If mixed methods) Is appropriate consideration given to the limitations associated with this integration, e.g., the divergence of qualitative and quantitative data (or results) in a triangulation design?	N/A	N/A	N/A	N/A	N/A	No	N/A	N/A	N/A	No
Average quality score:	1.3	1.7	0.8	1.7	0.8	1.4	1.2	1.25	0.9	1.8

Note. 'No' = clearly did not meet criterion (0 points), 'Can't tell' = insufficient information in the paper to decide or criterion only partially met (1 point), 'Yes' = clearly met criterion (2 points). Questions relating to completeness of outcome data were addressed based on data available at the first post-intervention assessment, as a common, comparable timepoint across studies (and focus of meta-analysis). RCT = Randomised Controlled Trial.

Figures

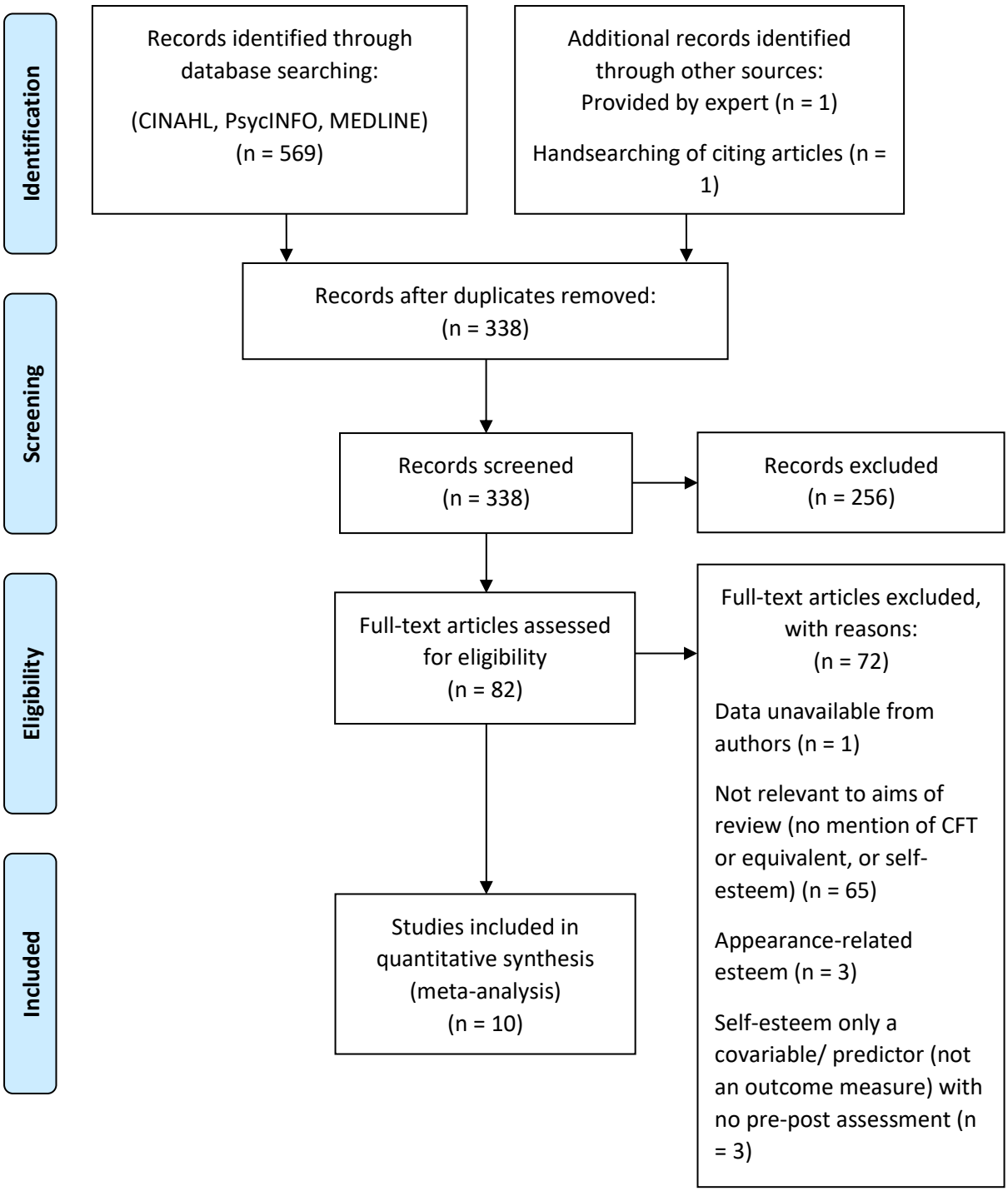

Figure 1. PRISMA flowchart of literature search

Study name	Hedges' g	CI Lower limit	CI Upper limit	Weight
[#1] Andersen & Rasmussen (2017)	0.75	0.51	0.98	13.82%
[#3] Gale et al (2014)	0.48	0.24	0.72	13.75%
[#4] Graser et al (2016)	0.02	-0.57	0.60	10.01%
[#6] Krieger et al (2019)	1.32	1.00	1.65	12.73%
[#7] Laithwaite et al (2009)	0.27	-0.16	0.70	11.57%
[#8] Lincoln et al (2013)	0.16	-0.14	0.47	12.99%
[#9] Murphy et al (2005)	0.36	0.18	0.54	14.41%
[#10] Rose et al (2018)	1.08	0.60	1.56	10.73%

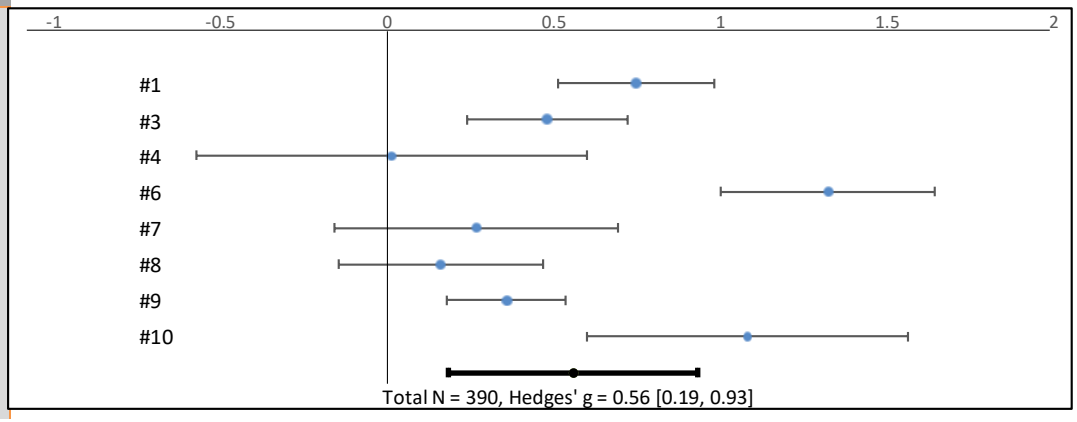

Figure 2. Forest plot of included studies

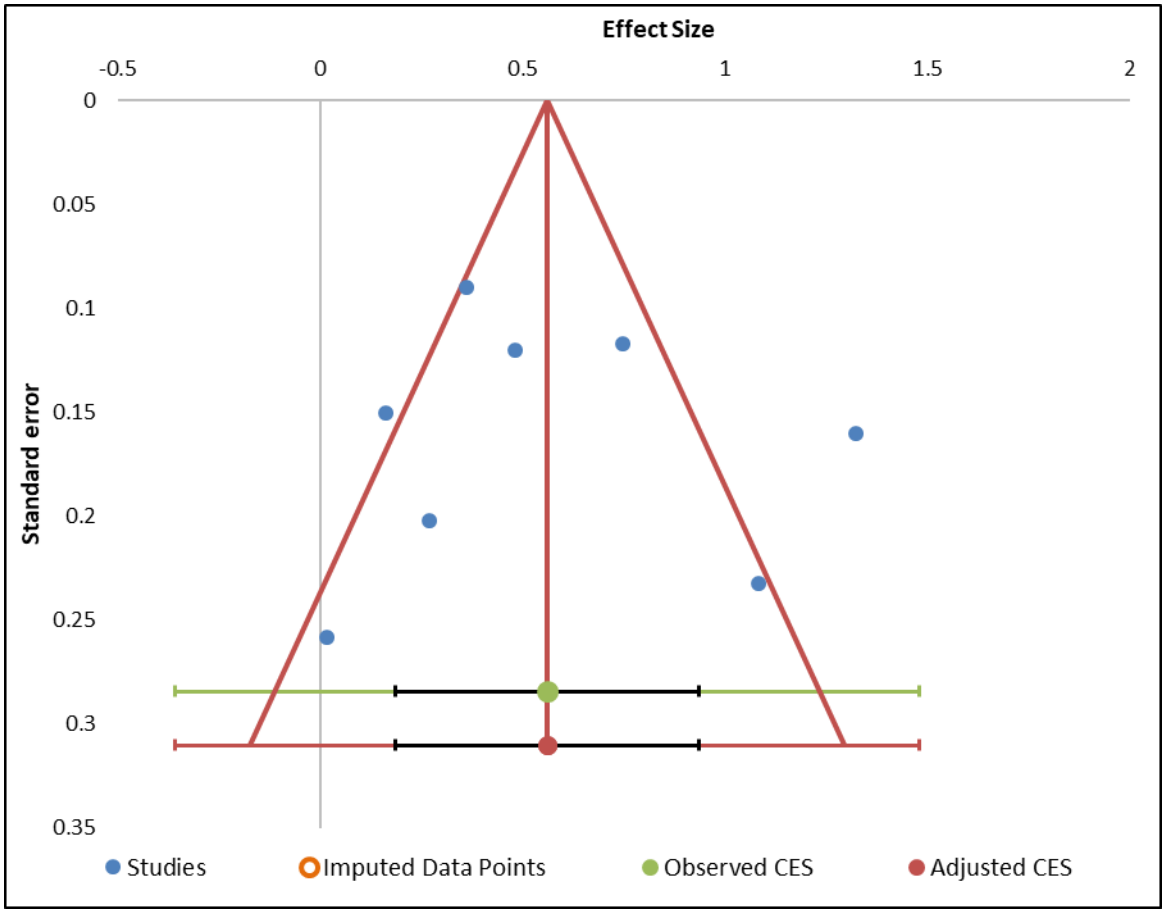

Figure 3. Funnel plot representing small-study or publication bias of included studies

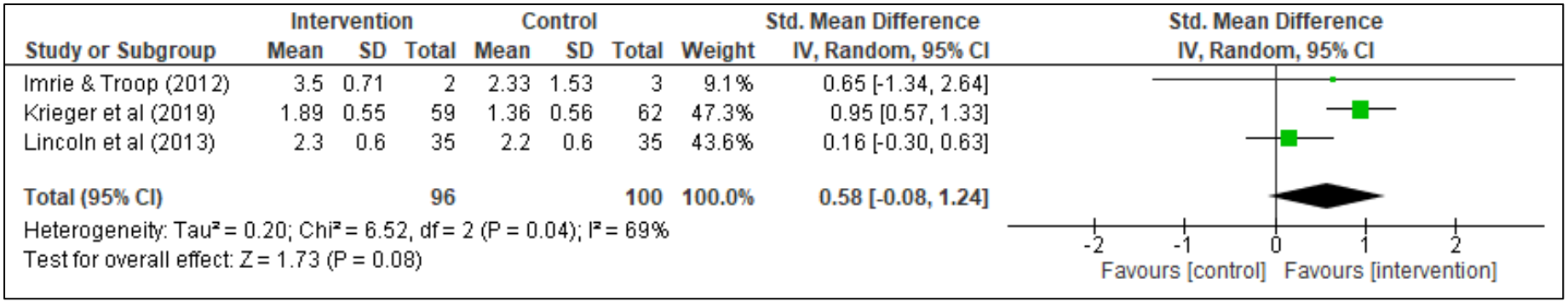

Figure 4. Forest plot of studies including a comparator group

Note. The Standard Mean Difference has been computed using Hedges' g.