Gratitude Interventions: Effective Self-Help? A Meta-Analysis of the Impact on Symptoms of Depression and Anxiety

David R. Cregg and Jennifer S. Cheavens

Department of Psychology, The Ohio State University, Columbus, OH, USA, 43210

In press at *Journal of Happiness Studies* © 2020, Springer Nature B.V.

This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available upon publication via its DOI: 10.1007/s10902-020-00236-6. You may also view the published typeset copy at: https://rdcu.be/b12vP

Corresponding author:
Jennifer S. Cheavens
147 Psychology Building
1835 Neil Avenue
Columbus, OH 43210
Phone: 614-247-6733
Email: cheavens.1@osu.edu
ORCID: 0000-0002-3778-9346

Conflict of Interest: The authors declare that they have no conflict of interest.

Acknowledgements: The authors would like to extend their sincerest thanks to Owen Morrish for his assistance with data extraction and the risk of bias assessment.

Abstract

Research suggests gratitude interventions, designed to increase appreciation of positive qualities, situations, and people in one's life, may improve psychological well-being (e.g., Seligman et al., 2005). Accordingly, mental health practitioners have promoted gratitude interventions as a means of self-help. However, results from previous reviews (Davis et al., 2016; Wood, Froh, & Geraghty, 2010) suggest that well-being improvements associated with gratitude interventions may be attributable to placebo effects. With this meta-analysis, we examined the efficacy of gratitude interventions ($k = 27$, $N = 3{,}675$) in reducing symptoms of depression and anxiety at post-test and follow-up periods. Gratitude interventions had a small effect on symptoms of depression and anxiety at both post-test ($g = -0.29$, $SE = 0.06$, $p < .01$) and follow-up ($g = -0.23$, $SE = 0.06$, $p < .01$). Correcting for attenuation from unreliability did not change results. Moderation analyses indicated effect sizes were larger for studies using waitlist, rather than active, control conditions at post-test and follow-up. We did not find consistent evidence for effects of other moderator variables (e.g., risk of bias, depressive symptom severity, or type of intervention used). Our results suggest the effects of gratitude interventions on symptoms of depression and anxiety are relatively modest. Therefore, we recommend individuals seeking to reduce symptoms of depression and anxiety engage in interventions with stronger evidence of efficacy for these symptoms.

Keywords: gratitude, thanks, intervention, positive psychology, depression, anxiety

Gratitude is a state of affirming the goodness or good things in one's life, accompanied by a recognition that the sources of this goodness lie at least partially outside the self, such as with the good intentions of another person (Emmons & Stern, 2013). Gratitude may be elicited by another person when he or she provides some aid or benefit, but it may also stem from non-social sources, such as a feeling of thanks for waking up in the morning (Wood, Froh, & Geraghty, 2010). A number of studies have demonstrated that gratitude has strong associations with measures of well-being, including positive correlations with positive affect, life satisfaction, extraversion, and forgiveness, and negative associations with neuroticism and substance abuse (see Watkins, 2014 and Wood et al., 2010 for reviews).

Additionally, the relationship between gratitude and symptoms of depression and anxiety has been the focus of a number of empirical studies (see Petrocchi & Couyoumdjian, 2016 for a review). For example, both Stoeckel, Weissbrod, and Ahrens (2014) and Watkins, Woodward, Stone, and Kolts (2003) found an inverse correlation of moderate strength between gratitude and symptoms of depression ($r = -.48$ and $r = -.56$, respectively). Likewise, Krumrei and Pargament (2008) reported an inverse association of moderate strength between gratitude and symptoms of anxiety ($r = -.46$). Additionally, Kendler et al. (2003) reported a reduced lifetime prevalence of Generalized Anxiety Disorder among those scoring higher in thankfulness, $OR = .82$. Although investigators are just beginning to uncover the specific mechanisms by which gratitude relates to depression and anxiety, there are several possible explanations for the inverse association between these variables. First, as Wood, Froh, and Geraghty (2010) noted, gratitude is associated with interpreting various stimuli and life events in positive terms, which contrasts with the selective attention to negative qualities of the self, the world, and the future that is characteristic of depression and anxiety (Mogg & Bradley, 2005; Peckham, McHugh, & Otto, 2010).

Consistent with Wood et al.'s idea, Petrocchi and Couyoumdjian (2016) found the inverse relationship between gratitude and symptoms of depression and anxiety is accounted for by a less critical, less punishing, and more compassionate view of oneself. Researchers have also found gratitude is associated with greater relationship connection and satisfaction (Algoe, Gable, & Maisel, 2010), which are well-established buffers against psychopathology (Seppala, Rossomando, & Doty, 2013). Finally, another way gratitude may guard against anxiety, in particular, is its relationship with uncertainty. A well-known characteristic of the worry observed in anxiety disorders is an intolerance of uncertainty (Carleton et al., 2012). Practicing gratitude may train one to be content in his or her present circumstances, whatever they may be, thus attenuating a fear of uncertain outcomes.

Given that gratitude is associated with a number of positive qualities, researchers have designed and tested several interventions to increase gratitude. For example, Emmons' and McCullough's (2003) seminal study asked participants to record five things for which they felt grateful each week for 10 weeks. They found significant increases in positive affect and hours spent exercising among the gratitude group, as well as improved sleep quality and reduced physical symptoms. Several researchers have also reported significant effects of gratitude interventions on symptoms of depression and anxiety. For instance, Seligman and colleagues (Seligman, Steen, Park, & Peterson, 2005) compared the effects of two gratitude interventions to a control condition (journaling about early memories) for reducing depressive symptoms. In the first gratitude condition, named "three good things," participants were instructed to keep a daily record of three good things and explain why they happened for one full week. In the second gratitude condition, named the "gratitude visit," participants were asked to write and personally deliver a letter to someone they had never properly thanked. Seligman et al. found the three good

things exercise led to reduced depressive symptoms for up to six months compared to the control group, and the gratitude visit reduced depressive symptoms for up to one month (although it had the largest positive improvements at post-test).

In a study on how gratitude interventions affect anxiety within a clinical sample, Kerr, O'Donovan, and Pepping (2015) recruited individuals on a waitlist for an outpatient psychology clinic. Client difficulties included depression, anxiety, and PTSD, as well as substance use and eating disorders. Individuals were randomized to either record up to five things they felt grateful for in the past day for 14 days, or to keep a daily mood diary in the control condition. The researchers found the gratitude intervention – but not the control task – significantly reduced anxiety over the 14-day period. Geraghty, Wood, and Hyland (2010) obtained similar results. Using an online community sample, they randomized participants to a waitlist condition or to a gratitude condition, in which participants listed six items for which they were grateful each day for 14 days. The gratitude condition led to significant reductions in worry, whereas the waitlist group experienced little change from baseline. Based on results such as these, researchers have suggested gratitude interventions may be an effective, low-cost, and easily implementable psychotherapeutic tool (Duckworth, Steen, & Seligman, 2005; Seligman, Rashid, & Parks, 2006).

However, in spite of these seemingly promising findings, a qualitative review by Wood, Froh, and Geraghty (2010) questioned the efficacy of gratitude interventions for a variety of outcomes, including depression and anxiety.[1] The authors argued many of the intervention studies included control groups that made inferences about the efficacy of gratitude interventions ambiguous. For example, they pointed out studies that, instead of including a neutral control

[1] Additional outcomes reviewed included body dissatisfaction, changes in positive and negative affect, life satisfaction, sleep quality, and physical pain.

task, included comparisons that may actually have harmful effects, such as listing hassles (Emmons & McCullough, 2003) or things one was unable to accomplish over the summer (Watkins et al., 2003). Relying on control conditions that may have effects in the opposite direction of the gratitude intervention (i.e., increasing symptoms of distress or reducing well-being) may inflate the effect size estimate. Conversely, a study by Lyubomirsky and colleagues with a more neutral control condition (writing about one's weekly schedule) found the gratitude intervention to have little effect on symptoms of depression (Lyubomirsky, Dickerhoof, Boehm, & Sheldon, 2011). Given the heterogeneity of these control groups, Wood, Froh, and Geraghty (2010) cautioned against a premature declaration of the success of gratitude interventions until more rigorous investigations could be conducted. Indeed, they noted of the 12 interventions included in their review, "only a very small number show that gratitude interventions are more effective than genuine controls" (Wood et al., 2010, p. 898).

Responding to Wood et al.'s (2010) critique, Davis et al. (2016) recently conducted a meta-analysis investigating the effects of gratitude interventions on gratitude, anxiety, and psychological well-being. They found gratitude interventions had generally limited effects, with effect sizes ranging from $d = 0.31$ for a measurement-only control (i.e., waitlist) to $d = -0.03$ for a "psychologically active" comparison (i.e., one that might be reasonably expected to promote psychological well-being, such as an automatic thought record). Based on these results, the authors concluded the evidence for the efficacy of gratitude interventions on psychological well-being, anxiety, and even gratitude itself is weak.

Although the conclusions reached by Wood et al. (2010), and later by Davis et al. (2016), question the benefits of gratitude interventions, these exercises have grown quite prominent in popular culture as a means of self-help. Paid-subscription smart-phone applications like Happify

(2016), best-selling books like *The Gratitude Diaries* (Kaplan, 2015), and magazine editorials (Graff, 2016) all claim that gratitude interventions will enhance one's life. Additionally, university wellness centers have begun to advocate for the use of gratitude interventions as therapeutic tools (Emmons, 2013; "Heart Centered Practices"). With such widespread use, it is important to determine whether gratitude interventions are indeed efficacious for specific psychological symptoms, such as depression and anxiety, which are among the most common mental health problems (Kendrick & Pilling, 2012).

Our meta-analysis extends the work of past reviews in several ways. First, new studies on gratitude interventions have been published since the prior reviews in this area. Our meta-analysis updates the literature for all studies published before May 17th, 2018 (see methods). Second, past meta-analyses of gratitude interventions have focused on positive psychological functioning, such as positive affect (Sin & Lyubomirsky, 2009) and life satisfaction (Davis et al., 2016). To date, ours is the first to conduct focused analyses on effects for symptoms of depression and anxiety. Sin and Lyubomirsky's (2009) meta-analysis assessed the effects of positive psychology interventions (PPIs) on symptoms of depression, but only four of the included studies used a gratitude intervention; thus, it was not possible to disentangle the effect of PPIs in general from the specific effect of gratitude interventions. Davis et al. (2016) also included measures of depression and anxiety. However, depression and life satisfaction scores were aggregated to create a "psychological well-being" outcome, and only 10 of the 21 studies for this aggregate outcome included a depression measure; thus, the unique effect for symptoms of depression could not be obtained. Their anxiety effect size also included a measure of marital satisfaction (Snyder, 1998), which may capture marital distress rather than anxiety-specific symptoms, such as worry (Meyer, Miller, Metzger, & Borkovec, 1990). Additionally, their

inclusion of comparison groups with demonstrated therapeutic value makes conclusions drawn from their anxiety analysis ambiguous. For example, evidence suggests that comparison conditions such as progressive muscle relaxation (Cheung, Molassiotis, & Chang, 2003) and automatic thought records (Persons & Burns, 1985) reduce symptoms of depression and anxiety. Therefore, in Davis et al.'s meta-analysis, it is uncertain whether gratitude interventions are ineffective for anxiety symptoms, or if they only improve symptoms to the same degree as other therapeutic interventions. Similarly, they did not distinguish between neutral and therapeutic controls in their moderator analyses of psychological well-being. As articulated by Chambless and Hollon (1998), such comparisons are inherently difficult to interpret, as null results may stem from a lack of power rather than equivalency between groups. Therefore, to eliminate this ambiguity, we only included studies with no-treatment (waitlist) and neutral comparison groups, i.e., active control tasks that were not intended to be therapeutic interventions and did not have empirical evidence of benefit for depression and/or anxiety.

Most recently, Dickens (2017) published a meta-analysis that found gratitude interventions have a small ($d = 0.13$) effect on depression when compared to waitlist or active control tasks. However, she did not report specific effect sizes for each of these two control types, but combined both under the label of "neutral conditions". Additionally, she applied a number of exclusion criteria that limit the generalizability of her results, i.e., excluding studies involving daily gratitude journals, studies lasting three days or less, and studies that involved multiple gratitude interventions. In the current meta-analysis, we did not apply such exclusions, but instead included study duration and type of intervention as moderators of the effect size. For studies with multiple gratitude interventions, we aggregated the effect sizes and ran our models with and without these aggregate scores to include as much information as possible (see details

under methods). With this broader inclusion criteria, we were able to include the largest number of studies examining depressive symptoms of any published meta-analysis to date. Additionally, the Dickens (2017) meta-analysis did not include symptoms of anxiety.

Additionally, neither Davis et al. (2016) nor Dickens (2017) performed a risk-of-bias assessment[2]. Such an assessment can reveal how study features such as participants' awareness of condition, dropout, and baseline differences may influence effect size estimates. For example, Bolier et al. (2013) found a larger effect size for PPIs with a greater risk of bias. Therefore, to assess for influences of bias, we conducted a risk-of-bias assessment using guidelines developed by the Cochrane Collaboration (Higgins, Altman, & Sterne, 2017).

Finally, for any variable assessed in a meta-analysis, there will be sources of error affecting its accuracy, such as measurement error. The influence of measurement error is particularly important for the outcome measure(s). For any given study in a meta-analysis, the effect size estimate will be attenuated to the extent that there is measurement error (unreliability) in the outcome measure(s) (Hedges & Olkin, 1985). If measurement error is present across studies, then the overall effect size will be attenuated by the cumulative impact of the unreliability across studies. Past meta-analyses of gratitude interventions have not controlled for measurement error in the outcome(s). Therefore, we conducted a correction for attenuation to minimize the influence of unreliability on our effect size estimates. The method for this correction is described in the methods section below.

With this meta-analysis, we examined the effects of gratitude interventions on symptoms of depression and anxiety at both immediate post-test and follow-up periods. Additionally, we

[2] Following the recommendation of the Cochrane Collaboration (Higgins et al., 2017), we use the term "risk of bias" here rather than "study quality," as a study may be of the highest possible quality, yet still contain important sources of bias.

assessed the influence of several moderator variables using meta-regression. Specifically, we were interested in determining: 1) the effect size of gratitude interventions on symptoms of depression and anxiety, considered separately; 2) the overall aggregate effect size on symptoms of depression and anxiety, considered together; and 3) whether the type of control group and other study characteristics (e.g., risk of bias, duration of intervention, type of intervention) moderated the effects.

Methods

Literature Search

We searched four databases for studies investigating the effects of gratitude interventions on symptoms of depression and anxiety (Cochrane Libraries, PsycINFO, PubMed, and Web of Science). Additionally, to maximize the number of potential studies included, we manually searched the reference sections of published review articles that discussed gratitude interventions (Carl, Soskin, Kerns, & Barlow, 2013; Davis et al., 2016; Sin & Lyubomirsky, 2009; Wood et al., 2010). Initial searches were conducted between June 1 – 2, 2016. A second search was conducted between May 17 – 18, 2018 to update the literature. See Online Resource 1 for detailed keyword profiles and filters applied to each database.

Inclusion Criteria

We used the following inclusion criteria:

1. Study was a scientific article in a relevant field and topic. Excluded studies were non-scientific essays from the humanities or scientific studies not focused on gratitude, such as miscellaneous biology or medical research.
2. Study used an experimental design with random assignment and a waitlist (measurement-only control) or neutral control condition. Excluded studies involved correlational or

qualitative research, or solely included comparisons with treatments of demonstrated efficacy for anxiety and/or depression (i.e., solely including other active treatments without a neutral control). A comparison activity was considered efficacious if it was intended to be therapeutic by the study authors, and it had empirical evidence demonstrating some benefit for depression and/or anxiety symptoms in past research (e.g., automatic thought records; Persons & Burns, 1985). If a study included both an active treatment and a neutral/waitlist control, we used the neutral control as the comparison group.

3. Adequate statistical information was available to compute an effect size. If the study did not contain adequate data, we contacted the corresponding or first author to retrieve the necessary information. If he or she could or would not provide it, we excluded that study.
4. Study included at least one measure of symptoms of either anxiety or depression. Excluded studies exclusively measured some other outcome(s), such as life satisfaction or physical health. A full-text review was conducted for all studies excluded for this reason to ensure that no depression or anxiety measures were reported in the manuscript.
5. Intervention was designed to induce or increase gratitude. Excluded studies used a non-gratitude intervention (such as mindfulness) or combined gratitude with another technique into a single group (such as a gratitude and best-possible-self exercise conducted together).

In total, we screened 1,277 abstracts for inclusion (953 unique studies across databases after removing duplicates). See Figure 1 for a flowchart of the screening process.

Data Extraction

Data extraction and a risk-of-bias assessment were conducted by the first author (DC) and independently checked by a second research assistant. If there was a discrepancy, each reviewer returned to the original paper to double check the data extraction. Disagreements were resolved by discussion. Data were collected on the study design, control group, intervention type, study duration, participant characteristics, baseline depressive symptoms, outcomes measured, post-test and follow-up raw data, sample size[3], publication status and year, the presence of a compliance or adherence check, and the risk-of-bias criteria outlined below. Raw data for each study (means, *SD*s, and reliability coefficients) can be found in Online Resource 2. For studies containing a depression measure with a published threshold score for clinically-relevant symptoms of depression, we also coded whether the sample's baseline depressive symptoms met the recommended threshold. See Online Resource 4 for a list of these threshold scores.

Additionally, we extracted reliability (Cronbach's alpha) coefficients at baseline for each outcome within each study in order to correct effect sizes for attenuation. If the specific timepoint was not reported (e.g., "reliabilities ranged from .91 to .93 across timepoints"), then we used the median value. If the reliability was not reported for each subscale of an outcome (e.g., the DASS-21), we used the reliability for the total scale. If the alpha value was not reported, we used the reliability coefficient from the original publication of the scale. For one study (Smullen, 2012), the reliability coefficient was not reported in either the study or the original publication of the scale. Therefore, we used the mean reliability value for all depression measures in our meta-analysis (0.86).

[3] If authors did not report sample sizes for individual groups, we assumed equal sample sizes by dividing the combined study sample by the number of groups in the study

We used seven categories for the risk-of-bias assessment based on criteria from the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins et al., 2017) and a previous meta-analysis of PPIs by Bolier et al. (2013). Each criterion was coded as 1 (meets criterion), 0 (fails to meet criterion), or N/A (insufficient information to determine criterion). A composite summary score was calculated for each study by adding the number of criteria met. We also created categorical groupings of bias risk. A study was categorized as low risk if 5 – 7 criteria were met, medium risk if 3 – 4 criteria were met, and high risk if 0 – 2 criteria were met. Both the summary score and categorical groupings were included as moderators in the meta-regression analyses. Additionally, because creating summary scores carries the limitation of assigning equal weight to all criteria (Higgins et al., 2017), we also analyzed each bias criterion as an independent categorical moderator in meta-regression, using the low-risk studies as the reference group.

The seven risk-of-bias criteria were: 1) Random sequence generation: did study authors describe a method for ensuring random assignment to groups; 2) Randomization concealment: were investigators unaware of assignment to groups and/or the randomization sequence; 3) Participants kept unaware of study condition; 4) Baseline comparability: were baseline values of depression and/or anxiety equivalent between groups, or were appropriate adjustments made to correct for baseline differences (e.g., including baseline depression as a covariate); 5) Participant attrition: if attrition occurred, was the attrition rate reported and analyzed, was attrition less than 50% of the initial randomized sample at post-test and follow-up, and was the attrition rate comparable between groups (no more than a 10% difference); 6) Correcting for missing data: if data were missing, did investigators make an attempt to correct for the missing data using an intention-to-treat analysis or other means of imputing data; 7) Miscellaneous: any idiosyncratic

risk of bias not captured in the other categories (e.g., excluding participants with psychiatric conditions or reporting a deviation from the pre-specified study protocol).

Statistical Analyses

We performed the primary calculations and analyses using the meta-analysis software OpenMEE (Dietz et al., 2015). We computed a standardized mean difference for each study using the formula for Hedges' *g* (Hedges & Olkin, 1985). Hedges' *g* is a form of Cohen's *d* that corrects for bias in sample size, with the same interpretation guidelines for small, medium, and large effects of $g = |0.2|$, $g = |0.5|$, and $g = |0.8^{+}|$, respectively. Values less than |0.2| can be considered trivial (Cohen, 1988). The formula for Hedges' *g* is:

$$g = \frac{\bar{\Delta}_{exp} - \bar{\Delta}_{ctl}}{\sqrt{\frac{(n_{exp-1})SD_{exp}^2 + (n_{ctl-1})SD_{ctl}^2}{(n_{total-2})}}} \times \left(1 - \frac{3}{4(n_{exp} + n_{ctl}) - 9}\right)$$

, where "exp" denotes the experimental group and "ctl" denotes the control group. The variance for Hedges' *g* is calculated by the formula: $V_g = J^2 \times V_d$. In this equation, V_g is the variance of *g*, *J* is a correction factor defined by $\left(1 - \frac{3}{4(n_{exp} + n_{ctl} - 2) - 1}\right)$, and V_d is the variance for Cohen's *d* defined by $\left(\frac{n_{exp} + n_{ctl}}{n_{exp} \times n_{ctl}} + \frac{d^2}{2(n_{exp} + n_{ctl})}\right)$. For studies using a pre/post-test design, we calculated a change score for each group by subtracting the pre-test mean from the post-test mean, thus obtaining a negative value if a reduction in symptoms occurred. If pre-test measures were not reported, we calculated the effect size using only post-test means. Only four studies did not include pre-test data (see Table 4). However, work by McKenzie, Herbison, and Deeks (2016) has demonstrated that mixing change scores with post-test data in meta-analyses gives an unbiased estimate of the effect size when heterogeneity is present and a random effects model is used, as was the case for our study.

For follow-up data, we subtracted the pre-test mean from the follow-up mean[4]. If multiple follow-up assessments were reported, we used the timepoint closest to one month, as this was the most common design across studies. Following the suggestion of Becker (1988), where possible we used the pre-test standard deviation to control for any intervention or practice effects that might affect the post-test variance. For studies reporting post-test only data, we used the post-test standard deviation. For studies reporting an intention-to-treat analysis, we used the means and standard deviations from the intention-to-treat data. If the intention-to-treat data were not reported for all groups (i.e., Southwell & Gould, 2017), we used the data from study completers. All studies were weighted by the inverse of their variance, with studies of smaller variance receiving a greater weight in analyses (see Table 2 for weights of all studies for the overall meta-analysis at post-test).

We conducted separate meta-analyses for depression and anxiety, as well as an overall aggregated analysis including both outcomes. For studies that assessed anxiety and depressive symptoms within the same sample, we treated each outcome as independent for the separate depression and anxiety analyses. However, for the overall analysis we aggregated their effect sizes and variances into a single outcome using the method described by Borenstein, Hedges, Higgins, and Rothstein (2009). This technique ensured we did not double-count these studies. We assumed a correlation of .65 between depression and anxiety scales based on average correlations reported in past research (Dobson, 1985). Finally, for studies that included multiple independent groups of gratitude interventions compared to the same control group, we combined

[4] One study included follow-up data without reporting pre-test data (Ozimkowski, 2007). Thus, we did not include follow-up data for this study.

the means and *SD*s across intervention groups into a single effect size for that study, again using the method outlined in Borenstein et al. (2009)[5].

Presumably, gratitude interventions have differing effect sizes based on unique study characteristics such as the duration and type of intervention. Therefore, we decided to use the more conservative random-effects model in our meta-analysis. The random effects model accounts for heterogeneity in effect sizes across studies when calculating a pooled effect. We used the DerSimonian and Laird (1986) estimator to adjust for this heterogeneity in our analyses.

Corrections for attenuation.

If the reliability coefficient of the outcome measure is known for each study, then the individual effect sizes can be corrected for attenuation due to unreliability prior to conducting the meta-analysis. These corrections will then provide an estimate of the disattenuated population effect size. Following the procedure outlined in Hedges and Olkin (1985), we corrected individual effect sizes by dividing Hedges' *g* by the square root of the Cronbach's alpha value. We then corrected the corresponding variance for each study by dividing the variance by the Cronbach's alpha value. We report the corrected (disattenuated) effect size estimates in the results section below. Additionally, corrected estimates for moderator analyses can be found in Online Resource 5.

Description of Moderators

We included several moderator variables for meta-regression analyses. Reference groups for categorical moderators are listed in Table 4. Consistent with the suggestion of Wood, Froh, and Geraghty (2010), we coded control groups as either waitlist or active controls. Participants in waitlist control groups completed no activities other than submitting symptom measures.

[5] We ran our analyses with and without these studies ($k = 2$). Results changed by only two-hundredths of a decimal point. Therefore, we retained the aggregated studies in all final analyses.

Participants in active control groups completed non-gratitude tasks matched to the gratitude interventions in terms of time. If a study contained a waitlist and an active control group, we selected the active control as the comparison. If there were two active control groups, we selected the more neutral of the two (e.g., using early childhood memories rather than early *positive* childhood memories; Mongrain & Anselmo-Matthews, 2012). We predicted active controls would have a smaller effect size than waitlist controls, as performing some structured activity may confer a greater expectation of benefit (placebo effect) than just completing measures in a waitlist group.

We coded the intervention type according to whether it was primarily interpersonal or intrapersonal in nature, or a combination of the two, following the suggestion made by Davis et al. (2016). Interpersonal interventions involved written and/or verbal expressions of gratitude to another person, such as the gratitude visit. Intrapersonal interventions involved personal reflections on things one has to be thankful for in life, but without instructions to direct the expression of gratitude toward a particular individual (such as gratitude journals and guided gratitude meditations). Combined interventions had participants complete both types of activities in a single group. We expected combined and interpersonal interventions to have larger effects than intrapersonal interventions at post-test (based on results from Seligman et al., 2005). If studies included multiple gratitude intervention groups, such as a separate interpersonal and intrapersonal condition, we excluded them from the moderator analysis for intervention type.

Additional planned moderators for the meta-regression were: 1) Online implementation: whether the study was conducted online or offline (i.e., in-person); 2) Publication status: was the study published or unpublished (i.e., a dissertation or thesis); 3) Depressive symptoms threshold: for depression measures that have published interpretation guidelines, did the sample's average

baseline depressive symptoms meet the published thresholds for a clinically relevant level of depressive symptomatology; 4) Baseline CES-D: the sample's average baseline score on the Center for Epidemiological Studies Depression scale (the most commonly used depression instrument in our meta-analysis; Radloff, 1977); 5) Year of publication; 6) Percentage of female participants; 7) Mean age of participants; 8) Duration of the intervention period, defined by both number of weeks and the number of days on which an activity was actually performed; and 9) whether the researchers included some form of an adherence or compliance check.

Based on the results of previous research (Harbaugh & Vasey, 2014; Sin & Lyubomirsky, 2009), we expected a larger effect size for more depressed samples, i.e., those meeting the depressive symptoms threshold and with a higher baseline level of depressive symptoms. We also expected a larger effect among samples with older adults, again based on past research by Sin & Lyubomirsky (2009). Finally, we expected a larger effect for published articles, offline studies, interventions of longer duration, and studies that included an adherence check. We had no *a priori* hypotheses for year of publication or percentage of females. Furthermore, because risk of bias in studies can influence effect sizes in either direction (underestimating or overestimating effects; Higgins et al., 2017), we did not specify directional hypotheses for the risk-of-bias analyses.

We assessed the influence of moderator variables using OpenMEE's meta-regression feature. For all moderator analyses, we used the overall aggregate effect size ($k = 27$) to maximize power. The only exceptions were depressive symptoms threshold and baseline CES-D, for which the depression-specific effect size was used.

Outliers

We identified potential outliers by examining studies for which their 95% confidence intervals laid entirely outside of the pooled confidence interval for all studies for each outcome (depression, anxiety, and the overall analysis). The statistical analyses were then repeated with the outliers excluded. As a double check on the identification of outliers, we also used the "influence" procedure contained within the R package "metafor" (Viechtbauer, 2010). The influence procedure computes various diagnostic tests to identify outliers by multiple criteria (e.g., Cook's distance and the influence of each study on the heterogeneity of variance). The influence procedure exactly replicated the results reported below, i.e., it identified the same studies as outliers with no additional outliers identified.

Estimates of Publication Bias

In addition to the risk-of-bias assessment, we used Rosenthal's (1979) fail-safe N to assess the potential for publication bias. The fail-safe N calculates the number of studies with null results that would be needed to inflate the observed *p* value above a specified alpha level in which the effect is no longer statistically significant. It can be conceived of as an assessment of the "file-drawer problem;" that is, the likelihood for unpublished studies with non-significant results to be excluded from the meta-analysis. If a large number of such studies potentially exist, the overall effect size is likely an overestimate of the true effect. A disadvantage of this method is its exclusive focus on *p* values, which only indirectly account for the effect size that is of primary interest. However, this method has clearly defined guidelines for interpretation and is suitable for meta-analyses of any size. Rosenthal (1979) suggests a fail-safe N value above $5k + 10$ reflects results that are tolerant to contradicting studies, where k is the number of studies included in the meta-analysis. Rosenthal noted this is a conservative threshold, meaning that if

the fail-safe N is well above this value, there is increased confidence that the observed effect size estimate is trustworthy.

We did not include a funnel plot test in our study, as funnel plots do not provide valid estimates of publication bias when fewer than 30 studies are included (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006). Similarly, due to the highly subjective nature of interpreting these plots, the Cochrane Collaboration has advised caution in their use (Sterne, Egger, Moher, & Boutron, 2017). However, interested readers may find the funnel plots for the overall meta-analysis at post-test and follow-up in Online Resource 3. Instead of funnel plots, we conducted a cumulative meta-analysis for the overall effect size at post-test and follow-up. For the cumulative meta-analyses, we followed the procedure described in chapter 13 of Schmidt and Hunter (2015), in which studies are sequentially added to the meta-analysis in order from largest to smallest sample size (i.e., the most precise studies to the least precise). Publication bias is evident if adding the smaller-sample studies causes an increase in the effect size's magnitude. Forest plots for the cumulative meta-analyses are also included in Online Resource 3.

Results

Reliability of Data Extraction

We assessed inter-rater reliability for the extracted data (prior to any discussion of discrepancies) using intra-class correlations (ICCs) for continuous variables and kappas for categorical variables. Reliability for the continuous variables was excellent, with ICCs ranging from .91 – 1.00 across post-test and follow-up periods, with a mean of .99. Reliability for the categorical variables other than the risk-of-bias assessment was likewise good, with kappas ranging from .70 – 1.00 with a mean of .91. Reliability for the risk-of-bias variables ranged from moderate to good, with kappas ranging from .42 – .76 with a mean of .63.

Study Attributes

Sample characteristics.

Full details of study characteristics are presented in Table 1. Twenty-seven studies were included in the meta-analysis, with a grand total of 3,675 participants at post-test and 2,318 participants at follow-up. Of the total participants at post-test, 2,030 participants were in the experimental (gratitude) group, and 1,645 were in the control group. The combined sample size for individual studies ranged from 22 to 514. Eighteen studies included an active control group (totaling 1,206 participants) and nine studies included a waitlist control (totaling 439 participants). Except for Ozimkowski (2007), all studies included a majority female sample, with the total percentage of females ranging from 55 – 100%. Mean age ranged from 19 – 69, with an average age of 32 years across studies. Thirteen studies (48% of the sample) included some form of an adherence or compliance check.

Twenty-one studies included only a depression measure, two studies included only an anxiety measure, and four studies included measures of both depression and anxiety. Only two studies included a clinical sample (participants with a diagnosed disorder or seeking treatment for a psychological condition; Kerr et al., 2015; Southwell & Gould, 2017). However, of the 18 studies for which published interpretation guidelines were available, 13 (over half of the studies with depression data) met the recommended threshold for a clinically relevant level of depressive symptoms, whereas only five did not meet this criteria. For the 10 studies that reported baseline CES-D data, average scores ranged from 13.53 to 34.15. The average CES-D score across studies was 20.31, which is 4.3 points above the recommended threshold of 16 for a clinically relevant level of depressive symptoms (Lewinsohn, Seeley, Roberts, & Allen, 1997).

The overall level of reliability (Cronbach's alpha) for outcomes was high. Reliability estimates for items on depression measures ranged from .60 – .94 with a mean of .86 (*SD* = .08). Reliability estimates for items on anxiety measures ranged from .80 – .95 with a mean of .88 (*SD* = .06).

Intervention characteristics.

The majority of studies used a pre/post-test design, with only four studies reporting post-test only data. Thirteen studies reported follow-up data, with the majority of studies (*k* = 9) using a one-month follow-up. Assessing duration by days (on which an activity was performed), the intervention period ranged in frequency from 1 day to 28 days across studies. Assessing duration by total weeks, the intervention period ranged from less than one week to eight weeks. The majority of studies (*k* = 19) used an intrapersonal gratitude intervention, four used an interpersonal intervention, two combined both types of interventions into a single condition, and two studies included both types of activities in separate groups and were thus not included in the moderator analysis. Twenty studies were conducted online and seven were conducted offline. The majority of studies (*k* = 22) were published articles, with only five unpublished theses or dissertations.

Risk of bias characteristics.

See Online Resource 2 for the full risk-of-bias assessment. The summary score ranged from 0 to 5, with an average of 3.07 (*SD* = 1.44). Six studies were categorized as low risk, 12 as medium risk, and nine as high risk. No study met all seven criteria. The majority of studies (21/27) met criteria for baseline comparability. About half of the studies met criteria for keeping participants unaware of the condition (14/27 studies). Less than half of the studies passed the criteria for attrition (10/27), sequence generation (9/27), or missing data (7/27). Only two studies

reported sufficient information about randomization concealment, with the majority of studies (19/27) categorized as N/A. Finally, only seven studies contained a miscellaneous risk of bias (e.g., excluding participants older than 45; Jackowska, Brown, Ronaldson, & Steptoe, 2016).

Post-test Main Effects

Main effects for post-test outcomes are presented in Table 3, and effect sizes for each study are listed in Table 2 (for the overall meta-analysis) and plotted in Figures 2 – 4. Gratitude interventions had a small but statistically significant effect on depressive symptoms, $k = 24$, $g = -0.23$, $SE = 0.05$, $p < .01$, $\tau^2 = 0.02$. The corrected (disattenuated) depression effect size was $g = -0.24$, $SE = 0.05$, $p < .01$, $\tau^2 = 0.03$. For anxiety ($k = 5$), gratitude interventions had a medium effect that approached statistical significance, $g = -0.52$, $SE = 0.30$, $p = .09$, $\tau^2 = 0.41$. The corrected anxiety effect size was $g = -0.55$, $SE = 0.32$, $p = .09$, $\tau^2 = 0.45$. For the overall meta-analysis ($k = 27$), gratitude interventions had a small, statistically significant effect on symptoms of depression and anxiety, $g = -0.29$, $SE = 0.06$, $p < .01$, $\tau^2 = 0.07$. The corrected overall effect size was $g = -0.31$, $SE = 0.07$, $p < .01$, $\tau^2 = 0.08$.

Tests for heterogeneity of effect sizes were significant for all outcomes (depression, anxiety, and the overall effect), suggesting significant variation in effect sizes across studies. For depression, effect sizes ranged from −0.96 to 0.07. For anxiety, effect sizes ranged from −1.64 to −0.02. For the overall analysis, effect sizes ranged from −1.64 to 0.07. We identified two outliers for post-test outcomes: Geraghty et al. (2010), with an anxiety effect size of $g = -1.64$, 95% CI [−2.08, −1.20]; and Ki (2009), with a depression effect size of $g = -0.96$, 95% CI [−1.28, −0.63]. These two studies were outliers for their respective outcomes as well as the overall meta-analysis (see Figures 2 – 4).

After removing these outliers from the dataset, the depression effect size was reduced to a trivial value, but remained statistically significant, $g = -0.17$, $SE = 0.04$, $p < .01$; this equates to a 26% reduction in magnitude of the effect size. The anxiety effect size was reduced by 69% to a trivial value and was statistically non-significant, $g = -0.16$, $SE = 0.11$, $p = .13$. For the overall meta-analysis, the effect size was trivial but remained statistically significant, $g = -0.18$, $SE = 0.04$, $p < .01$, equating to a 38% reduction in magnitude. Notably, after the removal of outliers, all heterogeneity tests became non-significant, suggesting the outliers accounted for the heterogeneity in effect sizes. Excluding these outliers for the corrected effect sizes resulted in the same findings: the corrected depression g was −0.18, anxiety $g = -0.17$, and the overall $g = -0.19$. Again, all heterogeneity tests became non-significant.

Follow-up Main Effects

Main effects for follow-up outcomes are presented in Table 3, and effect sizes for individual studies are plotted in Figure 5. Because only two studies reported anxiety follow-up data, we did not calculate an independent effect size for anxiety at follow-up. However, the available anxiety follow-up data were incorporated into the overall meta-analysis. Gratitude interventions had a small, statistically significant effect on depressive symptoms at follow-up, $k = 12$, $g = -0.24$, $SE = 0.06$, $p < .01$, $\tau^2 = 0.01$. The corrected depression effect size was $g = -0.25$, $SE = 0.06$, $p <.01$, $\tau^2 = 0.02$. For the overall meta-analysis ($k = 13$), gratitude interventions also had a small, statistically significant effect, $g = -0.23$, $SE = 0.06$, $p < .01$, $\tau^2 = 0.01$. The corrected overall effect size was $g = -0.24$, $SE = 0.06$, $p < .01$, $\tau^2 = 0.02$.

Heterogeneity tests were non-significant. However, a trend was observed for heterogeneity for depression, $\tau^2 = 0.01$, Q (11) = 17.19, $p = .10$. For both depression and the overall analysis, effect sizes ranged from −0.86 to 0.06. We identified one outlier for follow-up

outcomes: Cheng, Tsui, and Lam (2015) reported a depression effect size of $g = -0.86$, 95% CI [−1.35, −0.36].

After removing this outlier from the dataset, the depression effect size remained small and statistically significant ($p < .01$), and it was reduced to a value of $g = -0.20$, $SE = 0.05$, a 17% reduction in magnitude. For the overall meta-analysis, the effect size also remained small and statistically significant ($p < .01$), and it was reduced to a value of $g = -0.19$, $SE = 0.05$, a 17% reduction in magnitude. After removal of Cheng et al. (2015), the trend toward a significant heterogeneity test for depression was eliminated. Excluding this outlier for the corrected effect sizes resulted in the same findings: the corrected depression g was −0.21 and the overall $g = -0.20$. Again, the trend toward a significant heterogeneity test was eliminated.

Meta-Regression Analyses

Results of all meta-regression analyses are presented in Table 4. For both the post-test and follow-up time periods, the effect size was significantly moderated by type of control group. The pooled effect size was larger for studies with waitlist control groups than for those using active controls. No other moderators were significant for post-test or follow-up.

Due to the effect size differing based on the type of control group, we also tested interactions between all the continuous moderators and the type of control group used at post-test and follow-up. No significant interactions were found.

We repeated all moderator analyses with the corrected effect sizes. In all cases, the results were identical to the uncorrected analyses (see Online Resource 5).

Risk-of-bias assessment.

See Table 4 for results of the risk-of-bias analyses. Neither the summary bias score nor the categorical risk groupings (low, medium, and high risk) significantly moderated the effect

size at post-test or follow-up. When we examined individual risk categories, participants' awareness of condition significantly moderated the effect size at follow-up only. Studies judged to be at high risk of bias for threats to participants' awareness of condition ($k = 4$) had a larger pooled effect size ($g = -0.51$, $SE = 0.15$) than studies judged to be at low risk ($k = 7$; $g = -0.17$, $SE = 0.06$) and studies with insufficient information ($k = 2$; $g = -0.15$, $SE = 0.08$). No other individual risk categories had significant effects for either post-test or follow-up periods. Repeating analyses with the corrected effect sizes did not change the results.

Results of Publication Bias Estimates

Fail-safe N.

We computed the fail-safe N using an alpha of .05 for both the overall post-test and follow-up effect sizes. The post-test fail-safe N was 560, and the follow-up fail-safe N was 122. Both of these values are well above the $5k + 10$ guidelines of 145 and 75 for post-test and follow-up, respectively. Therefore, it is unlikely the overall effect sizes of −0.29 for post-test and −0.23 for follow-up are overestimations resulting from the file drawer problem.

Cumulative meta-analyses.

As can be observed in the forest plots in Online Resource 3, there was a trend for the effect size to increase with smaller samples at post-test. This result suggests that there is a possible bias toward publishing less precise studies (i.e., smaller samples) that report larger effects of gratitude interventions at post-test. There was no discernible pattern of bias at follow-up, though the fewer number of studies (13) limits the ability to detect visual patterns.

Discussion

The primary aim of our meta-analysis was to determine the effectiveness of gratitude interventions for symptoms of depression and anxiety. Considered altogether, our analyses

suggest gratitude interventions are of limited efficacy for reducing these symptoms. The overall effect sizes at post-test ($g = -0.29$) and follow-up ($g = -0.23$) suggest a small effect according to Cohen's (1988) guidelines.

The fail-safe N values also suggest these effect sizes are unlikely to be influenced by potential unpublished studies with null results. However, if such studies exist, they would only serve to diminish the effect sizes even further. This argument is supported by the cumulative meta-analyses, which suggests that if publication bias was present in our results, it was likely *over*estimating the effect size at post-test. Furthermore, excluding outliers reduced the effect sizes to still-smaller values, particularly at post-test.

Additionally, and as predicted, the effect size was smaller when gratitude interventions were compared to active control conditions. Consistent with past reviews (Davis et al., 2016; Lyubomirsky et al, 2009), we found gratitude interventions had a medium effect when compared with waitlist-only conditions, but only a trivial effect when compared with putatively inert control conditions involving any kind of activity.

Finally, it should be noted that the reliability distribution of the dependent variables was generally high. Effect-size estimates are only substantively attenuated when the level of reliability is low (Schmidt & Hunter, 2015), and we confirmed this in our dataset. In all cases, the analyses conducted with corrected effect sizes were consistent with the uncorrected estimates, i.e., results changed by only a few hundredths of a decimal point. Therefore, there is no evidence that the small effects obtained in this meta-analysis are a result of attenuation from unreliability. Based on our results, we agree with Davis et al. (2016) that a parsimonious explanation of gratitude interventions may be that they operate primarily through placebo effects, at least for depression and anxiety symptoms.

Excluding our prediction for the type of control group, we found little evidence to support our other hypotheses about moderator variables. Indeed, with one exception, none of the other tests of moderation were significant. The only exception was for participant awareness of condition, which significantly moderated effects at follow-up (but not at post-test). It is possible that among studies containing threats to participants' awareness of condition, researchers may have unwittingly influenced participants in ways that would favor gratitude interventions, such as communicating an expectation of benefit to the gratitude but not to the control group. However, given the lack of consistency between the post-test and follow-up effects for this criterion, we interpret these results very cautiously. We also advise caution in interpreting the null results of the other categorical moderators, particularly where there was an uneven number of studies in each category. The differential representation of studies between categories may create limited power to detect group differences. For example, only four studies included an interpersonal intervention, whereas 19 studies included an intrapersonal intervention. The substantially lower number of interpersonal interventions makes it difficult to draw conclusions about whether interpersonal interventions are truly equivalent to intrapersonal interventions in their effects. Indeed, evidence from Seligman et al. (2005) would suggest that they are not equivalent, as the interpersonal intervention had a more substantial impact on depression at post-test than the intrapersonal intervention in that study.

Limitations

Our results should be interpreted in light of several important limitations. First, and most crucially, the samples included in our meta-analysis were mostly comprised of unselected participants, i.e., individuals who were not selected based on severity of depressive or anxiety symptoms. Only two studies included clinical samples (Kerr et al., 2015; Southwell & Gould,

2017). Prior research suggests treatment effects may increase with depressive symptom severity. For example, Sin and Lyubomirsky (2009) found PPIs have a greater effect on reducing depressive symptoms among those who meet diagnostic status for a depressive disorder. Likewise, Harbaugh and Vasey (2014) found their gratitude intervention was effective only among those high in baseline depressive symptoms. Therefore, one objection to the results of our meta-analysis may be that the range of depressive symptoms was too restricted for gratitude interventions to have an effect. We think this explanation is unlikely for several reasons. First, over half of the studies in which symptoms of depression were assessed had a sample with baseline depressive symptoms meeting the recommended thresholds for clinically-relevant symptoms. We did not find evidence of a stronger effect for those samples in which participants, on average, met the established thresholds. Second, for studies including baseline CES-D data, the average scores ranged from 13.53 – 34.15. Six of the 10 studies reporting baseline CES-D data exceeded the threshold for clinically-relevant symptoms of depression (i.e., 16), with an average value across all studies of 20.31. However, we again did not find evidence that the depression effect size was moderated by baseline CES-D severity. Third, the post-test[6] depression effect sizes for the two studies with clinical samples were among some of the smallest in our meta-analysis (−0.12 and −0.06 for Kerr et al., 2015 and Southwell & Gould, 2017, respectively). Based on these considerations, it appears that a sufficient range of depressive symptoms was present in our meta-analysis, but the effect of gratitude interventions did not increase with greater depressive symptomatology. Consequently, it seems unlikely that range restriction in our study accounts for the small effect size of gratitude interventions.

[6] These studies did not include follow-up data.

That said, we certainly acknowledge that though depressive symptoms exist on a continuum, meeting a threshold for clinically relevant symptoms on a depression measure does not equate to a diagnosis of a depressive disorder made by a mental health professional (American Psychiatric Association, 2013). Therefore, our meta-analysis should be interpreted with this distinction between symptoms and diagnostic classes in mind. It is possible that future researchers may find a greater benefit of gratitude interventions for depressive symptoms by limiting recruitment to those meeting diagnostic status for a depressive disorder, which would allow for testing gratitude interventions in a more severely impaired population. However, to date, efforts along this line have yielded mixed results (Celano et al., 2017; Taylor, Lyubomirsky, & Stein, 2017).

Likewise, there may be other moderators that could influence gratitude interventions' effectiveness that we did not assess, such as one's level of self-criticism or emotional neediness (Sergeant & Mongrain, 2011). The instructions given to participants could also moderate effects. For example, previous research by Sheldon, Boehm, and Lyubomirsky (2012) suggests that variety is an important moderator of PPIs. Thus, if participants were instructed to list three *new* things they are grateful for each day, it could reduce some of the hedonic adaptation that may occur from repeating the same gratitude list daily. Additionally, it is possible that some participants perceived the instructions differently or performed the activity differently within the same condition of a study, i.e., there may be unreliability in the treatment variable (Schmidt & Hunter, 2015). However, the studies in our meta-analysis did not report an interrater reliability coefficient for the treatment variable, thus making it impossible to estimate the effect of treatment reliability on outcomes. Understanding whether instructions, treatment reliability, or other potential moderators increase the effectiveness of gratitude interventions is an important

direction for future research. That said, if we base our conclusions on the currently available data, there is little evidence to suggest gratitude interventions are efficacious for reducing symptoms of depression or anxiety. Accordingly, suggestions by researchers to use gratitude interventions as a psychotherapeutic tool (e.g., Emmons & Stern, 2013) should be taken cautiously until more substantial benefits can be demonstrated. Consequently, we recommend individuals seek more well-established treatments for difficulties with depression or anxiety symptoms until stronger benefits of gratitude interventions are found. For example, meta-analytic evidence suggests that computerized treatments for depression and anxiety, which are relatively low-cost and easily accessible, have strong effect sizes across comparison groups (Andrews, Cuijpers, Craske, McEvoy, & Titov, 2010).

A second limitation of our meta-analysis is the small number of studies ($k = 5$) for the anxiety effect size, which leads us to interpret this effect size cautiously. Although the effect size was of a medium magnitude ($g = -0.52$), it was statistically non-significant with a wide confidence interval ranging from a strong, beneficial effect to a weak, harmful effect (−1.11 to 0.08). Indeed, removing the outlier of Geraghty et al. (2010) eliminated the heterogeneity in the effect size and dropped its value to a trivial level of $g = -0.16$. Future investigations with measures of anxiety are needed so that meta-analytic estimates will be better powered and less influenced by outliers. Including more anxiety studies would also allow investigators to examine if the anxiety effect is moderated by symptom severity, as we were unable to examine this possibility with the small number of studies with anxiety measures currently available. Nevertheless, based on the current data, it appears gratitude interventions have limited efficacy for anxiety symptoms.

Third, this meta-analysis applies only to gratitude interventions' specific effects on symptoms of depression and anxiety as standalone interventions. It is not our intention to dismiss the value of gratitude interventions in general. It is entirely possible these interventions are more powerful for anxiety or depressive symptoms when they are integrated into a larger treatment package, as suggested by prior randomized trials with positive psychology exercises (Seligman, Rashid, & Parks, 2006; Taylor, Lyubomirsky, & Stein, 2017). Additionally, gratitude interventions may have stronger effects for constructs like relationship quality or general well-being, as Dickens' (2017) meta-analysis would suggest. However, it is important to understand the outcomes for which gratitude interventions are the most efficacious, and then recommend these interventions only when individuals seek to impact those particular outcomes. Indeed, our meta-analysis suggests that whatever the merits gratitude interventions have for other outcomes, they are not efficacious for symptoms of depression or anxiety as standalone interventions.

Fourth, there may be limitations for generalizability of our results based on other sample characteristics such as age and sex. Though we did not find evidence that age and sex moderated effects, all but one study included a majority female sample, and only five studies contained a sample with a mean age of 40 or above. Prior research suggests women (Kashdan, Mishra, Breen, & Froh, 2009) and older adults (Sin & Lyubomirsky, 2009) experience gratitude in unique ways. Therefore, it is unclear if our results generalize to samples with fewer women and a greater number of older adults.

Finally, our results only apply to gratitude *interventions*. They do not inform us about the association of gratitude as a general disposition with depression and anxiety. As we mentioned in the introduction, there is strong evidence that higher trait gratitude is associated with reduced psychopathology and greater well-being (Wood et al., 2010). Therefore, gratitude as a general

disposition may still be a vital factor in human flourishing, and the efficacy (or lack thereof) of gratitude interventions should in no way be interpreted to mean gratitude is not an important element of well-being and the good life.

Conclusion

Based on the currently available data, we find limited evidence for the efficacy of gratitude interventions in reducing symptoms of depression and anxiety. They have a medium-sized effect when compared to no intervention at all, but a small effect when compared with any active control task. Nevertheless, future investigators may discover individual differences that moderate the effectiveness of gratitude interventions, such as severity of psychopathology or qualities like self-criticism and emotional neediness. Such distinctions will be crucial to uncover, especially as exercises like gratitude journals have begun to permeate into popular culture as a means of self-help. We believe that until gratitude interventions are shown to be more powerful, the suggestions to use the existing approaches as tools for reducing symptoms of depression or anxiety should be considered with caution.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Algoe, S. B., Gable, S. L., & Maisel, N. C. (2010). It's the little things: Everyday gratitude as a booster shot for romantic relationships. *Personal Relationships, 17*, 217-233. doi:10.1111/j.1475-6811.2010.01273.x

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Andrews, G., Cuijpers, P., Craske, M. G., McEvoy, P., & Titov, N. (2010). Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: A meta-analysis. *Plos One, 5*, e13196.

Beck, A. T. (1976). *Cognitive therapy and the emotional disorders.* Oxford, England: International Universities Press.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257 - 278. doi:10.1111/j.2044-8317.1988.tb00901.x

Bolier, L., Haverman, M., Westerhof, G. J., Riper, H., Smit, F., & Bohlmeijer, E. (2013). Positive psychology interventions: A meta-analysis of randomized controlled studies. *BMC Public Health, 13*, 119. doi:10.1186/1471-2458-13-119

*Booker, J. A., & Dunsmore, J. C. (2017). Expressive writing and well-being during the transition to college: Comparison of emotion-disclosing and gratitude-focused writing. *Journal of Social and Clinical Psychology, 36*, 580-606. doi:10.1521/jscp.2017.36.7.580

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, United Kingdom: John Wiley & Sons, Ltd. doi:10.1002/9780470743386

Carl, J. R., Soskin, D. P., Kerns, C., & Barlow, D. H. (2013). Positive emotion regulation in emotional disorders: A theoretical review. *Clinical Psychology Review, 33*, 343-360.

Carleton, R. N., Mulvogue, M. K., Thibodeau, M. A., McCabe, R. E., Antony, M. M., & Asmundson, G. J. (2012). Increasingly certain about uncertainty: Intolerance of uncertainty across anxiety and depression. *Journal of Anxiety Disorders*, 26, 468-479. doi:10.1016/j.janxdis.2012.01.011

Celano, C. M., Beale, E. E., Mastromauro, C. A., Stewart, J. G., Millstein, R. A., Auerbach, R. P., Bedoya, C.A., & Huffman, J. C. (2017). Psychological interventions to reduce suicidality in high-risk patients with major depression: A randomized controlled trial. *Psychological Medicine, 47*, 810-821. doi:10.1017/S0033291716002798

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7-18. doi:10.1037/0022-006X.66.1.7

*Cheng, S. T., Tsui, P. K., & Lam, J. H. (2015). Improving mental health in health care practitioners: Randomized controlled trial of a gratitude intervention. *Journal of Consulting and Clinical Psychology, 83*, 177-186. doi:10.1037/a0037895

Cheung, Y. L., Molassiotis, A., & Chang, A. M. (2003). The effect of progressive muscle relaxation training on anxiety and quality of life after stoma surgery in colorectal cancer patients. *Psycho-Oncology, 12*, 254. doi:10.1002/pon.638

Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic

Davis, D. E., Choe, E., Meyers, J., Wade, N., Varjas, K., Gifford, A., Quinn, A., Hook, J., Tongeren, D.R., Griffin, B.J., Worthington, E. L. (2016). Thankful for the little things: A meta-analysis of gratitude interventions. *Journal of Counseling Psychology, 63*, 20-31. doi:10.1037/cou0000107

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*, 177-178.

Dietz, G., Dahabreh, I. J., Gurevitch, J., Lajeunesse, M. J., Schmid, C. H., Trikalinos, T. A. & Wallace, B. C. (2015). OpenMEE: Software for ecological and evolutionary meta-analysis.

Dobson, K. S. (1985). An analysis of anxiety and depression scales. *Journal of Personality Assessment, 49*, 522-527.

Duckworth, A. L., Steen, T. A., & Seligman, M. E. P. (2005). Positive psychology in clinical practice. *Annu Rev Clin Psychol, 1*, 629-651. doi:10.1146/annurev.clinpsy.1.102803.144154

Dickens, L. R. (2017). Using gratitude to promote positive change: A series of meta-analyses investigating the effectiveness of gratitude interventions. *Basic and Applied Social Psychology, 39*, 193-208. doi:10.1080/01973533.2017.1323638

Emmons, R. A. (2013). Gratitude practice explained. Retrieved from http://ei.yale.edu/what-is-gratitude/

Emmons, R. A., & McCullough, M. E. (2003). Counting blessings versus burdens: An experimental investigation of gratitude and subjective well-being in daily life. *Journal of Personality and Social Psychology, 84*, 377-389.

Emmons, R. A., & Stern, R. (2013). Gratitude as a psychotherapeutic intervention. *Journal of Clinical Psychology, 69*, 846-855. doi:10.1002/jclp.22020

*Gander, F., Proyer, R. T., Ruch, W., & Wyss, T. (2013). Strength-based positive interventions: Further evidence for their potential in enhancing well-being and alleviating depression. *Journal of Happiness Studies, 14*, 1241-1259. doi:10.1007/s10902-012-9380-0

*Gavian, M. E. (2011). The effects of relaxation and gratitude interventions on stress outcomes. *Dissertation Abstracts International: Section b: The Sciences and Engineering, 73*, 1248.

*Geraghty, A. W. A. (2010). *From placebo to self-help: Investigating retention, outcome and mechanisms in self-directed gratitude interventions* (Unpublished doctoral dissertation). University of Plymouth, Plymouth, Devon, United Kingdom.

*Geraghty, A. W. A., Wood, A. M., & Hyland, M. E. (2010). Attrition from self-directed interventions: Investigating the relationship between psychological predictors, intervention content and dropout from a body dissatisfaction intervention. *Social Science & Medicine, 71*, 30-37. doi:10.1016/j.socscimed.2010.03.007

Graff, M. (2016, December). The letter. *Southwest: The Magazine*, 82-91.

Happify. (2016). Happify, Inc. (Version 1.4.0R5433.bbbfa2d) [Mobile application software]. Retrieved from http://my.happify.com/

*Harbaugh, C. N., & Vasey, M. W. (2014). When do people benefit from gratitude practice? *Journal of Positive Psychology, 9*, 535-546. doi:10.1080/17439760.2014.927905

Heart centered practices. Retrieved from https://wexnermedical.osu.edu/integrative-complementary-medicine/heart-centered-practices

Hedges, L. V., & Olkin, L. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Higgins, J., Altman, D., & Sterne, J. (2017). Chapter 8: Assessing risk-of-bias in included studies. In J. Higgins, R. Churchill, J. Chandler & M. Cumpston (Eds.), *Cochrane handbook for systematic reviews of interventions* (version 5.2.0). Cochrane. Retrieved from www.training.cochrane.org/handbook

*Jackowska, M., Brown, J., Ronaldson, A., & Steptoe, A. (2016). The impact of a brief gratitude intervention on subjective well-being, biology and sleep. *Journal of Health Psychology, 21*, 2207-2217. doi:10.1177/1359105315572455

Kaplan, J. (2015). *The gratitude diaries: How a year looking on the bright side can transform your life*. New York, New York: Penguin Publishing Group.

Kashdan, T., Mishra, A., Breen , A., & Froh, J. (2009). Gender differences in gratitude: Examining appraisals, narratives, the willingness to express emotions, and changes in psychological needs. *J Pers, 77*, 691-730. doi:10.1111/j.1467-6494.2009.00562.x

Kendler, K. S., Liu, X. Q., Gardner, C. O., McCullough, M. E., Larson, D., & Prescott, C. A. (2003). Dimensions of religiosity and their relationship to lifetime psychiatric and substance use disorders. *The American Journal of Psychiatry, 160*, 496-503. doi:10.1176/appi.ajp.160.3.496

Kendrick, T., & Pilling, S. (2012). Common mental health disorders — identification and pathways to care: NICE clinical guideline. *Br J Gen Pract, 62*, 47-49. doi:10.3399/bjgp12X616481

*Kerr, S. L., O'Donovan, A., & Pepping, C. A. (2015). Can gratitude and kindness interventions enhance well-being in a clinical sample? *Journal of Happiness Studies, 16*, 17-36. doi:10.1007/s10902-013-9492-1

*Ki, T. P. (2009). *Gratitude and stress of health care professionals in hong kong* (Unpublished doctoral dissertation). City University of Hong Kong, Kowloon, Hong Kong, China.

Krumrei, E. J., & Pargament, K. I. (2008). *Are gratitude and spirituality protective factors against psychopathology?* Poster session presented at the 5th Biennial International Conference on Personal Meaning.

*Lambert, N. M., Fincham, F. D., & Stillman, T. F. (2012). Gratitude and depressive symptoms: The role of positive reframing and positive emotion. *Cognition & Emotion, 26*, 615-633. doi:10.1080/02699931.2011.595393

Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *BMJ: British Medical Journal, 333*, 597-600.

Lewinsohn, P.M., Seeley, J.R., Roberts, R.E., & Allen, N.B. (1997). Center for Epidemiological Studies-Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychology and Aging, 12*, 277- 287.

*Lyubomirsky, S., Dickerhoof, R., Boehm, J. K., & Sheldon, K. M. (2011). Becoming happier takes both a will and a proper way: An experimental longitudinal intervention to boost well-being. *Emotion, 11*, 391-402. doi:10.1037/a0022575

*Manthey, L., Vehreschild, V., & Renner, K. (2016). Effectiveness of two cognitive interventions promoting happiness with video-based online instructions. *Journal of Happiness Studies, 17*, 319-339. doi:10.1007/s10902-014-9596-2

McKenzie, J. E., Herbison, G. P., & Deeks, J. J. (2016). Impact of analysing continuous outcomes using final values, change scores and analysis of covariance on the performance of meta-analytic methods: A simulation study. *Research Synthesis Methods, 7*, 371-386. doi:10.1002/jrsm.1196

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy, 28*, 487-495. doi:0005-7967(90)90135-6

Mogg, K., & Bradley, B. P. (2005). Attentional bias in generalized anxiety disorder versus depressive disorder. *Cognitive Therapy and Research*, *29*, 29-45. doi:10.1007/s10608-005-1646-y

*Mongrain, M., & Anselmo-Matthews, T. (2012). Do positive psychology exercises work? A replication of seligman et al. *Journal of Clinical Psychology, 68*, 382-389. doi:10.1002/jclp.21839

*O'Leary, K., & Dockray, S. (2015). The effects of two novel gratitude and mindfulness interventions on well-being. *The Journal of Alternative and Complementary Medicine, 21*, 243-245. doi:10.1089/acm.2014.0119

*Ozimkowski, K. M. (2007). *The gratitude visit in children and adolescents: An investigation of gratitude and subjective well-being* (Unpublished doctoral dissertation). Hofstra University, New York, New York.

Peckham, A. D., McHugh, R. K., & Otto, M. W. (2010). A meta-analysis of the magnitude of biased attention in depression. *Depression and Anxiety, 27*, 1135-1142. doi:10.1002/da.20755

Persons, J. B., & Burns, D. D. (1985). Mechanisms of action of cognitive therapy: The relative contributions of technical and interpersonal interventions. *Cognitive Therapy and Research, 9*, 539-551. doi:10.1007/BF01173007

Petrocchi, N., & Couyoumdjian, A. (2016). The impact of gratitude on depression and anxiety: The mediating role of criticizing, attacking, and reassuring the self. *Self and Identity, 15*, 191-205. doi:10.1080/15298868.2015.1095794

*Proyer, R. T., Gander, F., Wellenzohn, S., & Ruch, W. (2014). Positive psychology interventions in people aged 50-79 years: Long-term effects of placebo-controlled online interventions on well-being and depression. *Aging & Mental Health, 18*, 997-1005. doi:10.1080/13607863.2014.899978

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

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638-641. doi:10.1037/0033-2909.86.3.638

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). London, United Kingdom: SAGE Publications, Inc. doi:10.4135/9781483398105

Seligman, M. E. P., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. *The American Psychologist, 60*, 410-421. doi:10.1037/0003-066X.60.5.410

Seligman, M. E. P., Rashid, T., & Parks, A. C. (2006). Positive psychotherapy. *American Psychologist, 61*, 774-788. doi:10.1037/0003-066X.61.8.774

*Senf, K., & Liau, A. K. (2012). The effects of positive interventions on happiness and depressive symptoms, with an examination of personality as a moderator. *Journal of Happiness Studies, 14*, 591-612. doi:10.1007/s10902-012-9344-4

Seppala, E., Rossomando, T., & Doty, J. R. (2013). Social connection and compassion: Important predictors of health and well-being. *Social Research, 80*, 411-430.

*Sergeant, S., & Mongrain, M. (2011). Are positive psychology exercises helpful for people with depressive personality styles? *Journal of Positive Psychology, 6*(4), 260-272. doi:10.1080/17439760.2011.577089

Sheldon, K. M., Boehm, J. K., & Lyubomirsky, S. (2012). Variety is the spice of happiness: The hedonic adaptation prevention (HAP) model. In I. Boniwell & S. David (Eds.), *Oxford handbook of happiness* (pp. 901–914). Oxford, England: Oxford University Press.

Sin, N. L., & Lyubomirsky, S. (2009). Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: A practice-friendly meta-analysis. *Journal of Clinical Psychology, 65*, 467-487. doi:10.1002/jclp.20593

*Smullen, A. M. (2012). *A gratitude intervention with older adults* (Unpublished master's thesis). Kean University, Union, New Jersey.

Snyder, D. (1998). Marital Satisfaction Inventory—Revised. Los Angeles, CA: Western Psychological Services.

*Southwell, S., & Gould, E. (2017). A randomised wait list-controlled pre–post–follow-up trial of a gratitude diary with a distressed sample. *The Journal of Positive Psychology, 12*, 579-593. doi:10.1080/17439760.2016.1221127

Sterne, J., Egger, M., Moher, D., & Boutron, I. (2017). Chapter 10: Addressing reporting biases. In J. Higgins, R. Churchill, J. Chandler & M. Cumpston (Eds.), *Cochrane handbook for systematic reviews of interventions* (version 5.2.0). Cochrane. Retrieved from www.training.cochrane.org/handbook

Stoeckel, M., Weissbrod, C., & Ahrens, A. (2014). The adolescent response to parental illness: The influence of dispositional gratitude. *Journal of Child and Family Studies, 24*, 1501–1509. Advance online publication. doi:10.1007/s10826-014-9955-y

Taylor C, Lyubomirsky S, & Stein, M. (2017). Upregulating the positive affect system in anxiety and depression: Outcomes of a positive activity intervention. *Depress Anxiety, 34*, 267–280. doi:10.1002/da.22593

*Timmons, L., & Ekas, N. V. (2018). Giving thanks: Findings from a gratitude intervention with mothers of children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 49*, 13-24. doi:10.1016/j.rasd.2018.01.008

*Toepfer, S., Cichy, K., & Peters, P. (2012). Letters of gratitude: Further evidence for author benefits. *Journal of Happiness Studies, 13*, 187-201. doi:10.1007/s10902-011-9257-7

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48.

Watkins, P. C. (2014). *Gratitude and the good life*. Springer Netherlands. doi:10.1007/978-94-007-7253-3

Watkins, P. C., Woodward, K., Stone, T., Kolts, R.L. (2003). Gratitude and happiness: Development of a measure of gratitude, and relationships with subjective well-being. *Social Behavior and Personality, 31*, 431-452.

*Watkins, P. C., Cruz, L., Holben, H., & Kolts, R. L. (2008). Taking care of business? Grateful processing of unpleasant memories. *The Journal of Positive Psychology, 3*, 87-99. doi:10.1080/17439760701760567

*Watkins, P. C., Uhder, J., & Pichinevskiy, S. (2015). Grateful recounting enhances subjective well-being: The importance of grateful processing. *Journal of Positive Psychology, 10*, 91-98. doi:10.1080/17439760.2014.927909

*Wolfe, W. L., & Patterson, K. (2017). Comparison of a gratitude-based and cognitive restructuring intervention for body dissatisfaction and dysfunctional eating behavior in college women. *Eating Disorders, 25*, 330-344. doi:10.1080/10640266.2017.1279908

Wood, A. M., Froh, J. J., & Geraghty, A. W. (2010). Gratitude and well-being: A review and theoretical integration. *Clinical Psychology Review, 30*, 890-905. doi:10.1016/j.cpr.2010.03.005

Table 1

Characteristics of Included Studies

Author	Year	N	Outcome(s)	Reliability (Alpha)[a]	Change Score or Post-test Only	Control	Intvn.	Intvn. Type	% Female[b]	Mean Age[c]	Pop.	Dep. Threshold	Baseline CES-D	Duration (Days, Weeks)	Follow-up	Online	Compliance
Booker & Dunsmore	2017	108	D: CES-D	.89	Change	WLC	Gratitude journal	Intra	70	19	NC (college)	Yes	31.82	4d, 1w	1m	Yes	Yes
Cheng et al.	2015	68	D: CES-D10	.74	Change	WLC	Gratitude list	Intra	55		NC (healthcare professionals)	No		8d, 4w	3m	Yes	Yes
Gander et al.[e]	2013	335	D: CES-D	.92	Change	Early memories	Gratitude visit and list		95	45	NC (community)	No	13.53		1m	Yes	Yes
Gavian (Unpublished)	2011	174	A, D: DASS-21	.92	Change	Daily schedule	Gratitude list	Intra	70		NC (college)	Yes		7d, 1w	1m	Yes	Yes
Geraghty et al.	2010	108	A: PSWQ	.91	Change	WLC	Gratitude list	Intra	86	37	NC (community)			14d, 2w		Yes	No
Geraghty Study 4 (Unpublished)	2010	149	D: PHQ-9	.88	Change	WLC	Gratitude list	Intra	81	37	NC (community)	Yes		14d, 2w		Yes	Yes
Harbaugh & Vasey	2014	108	D: CES-D	.92	Change	Daily events	Gratitude list	Intra	70	20	NC (college)	No	15.50	14d, 2w		Yes	Yes
Jackowska et al.	2016	81	A, D: HADS[f]	.84	Change	Daily events	Gratitude diary & letter	Comb	100		NC (college)			6d, 2w		No	Yes
Kerr et al.	2015	31	A, D: DASS-21	.80, .91	Change	Mood diary	Gratitude list	Intra	75	43	C (outpatient clients)	Yes		14d, 2w		No	No
Ki (Unpublished)	2009	161	D: CES-D10	.60	Change	Hassle list	Gratitude list	Intra	58		NC (healthcare professionals)	Yes		8d, 4w		Yes	No

Author	Year	N	Outcome(s)	Reliability (Alpha)[a]	Change Score or Post-test Only	Control	Intvn.	Intvn. Type	% Female[b]	Mean Age[c]	Pop.	Dep. Threshold	Baseline CES-D	Duration (Days, Weeks)	Follow-up	Online	Compliance
Lambert et al. study 5	2012	112	D: 4-Item State Depressive Symptoms Scale	.71	Post-test	Record things learned in human development class	Gratitude journal	Intra	88		NC (college)			1d, 0w		Yes	No
Lambert et al. study 7	2012	89	D: 4-Item State Depressive Symptoms Scale	.80	Post-test	Journal of insights from college courses	Gratitude journal	Intra	82		NC (college)			28d, 4w		Yes	No
Lyubomirsky et al.	2011	208	D: CES-D	.88	Change	Weekly events	Gratitude letter	Inter	71	20	NC (college)	Yes	34.15	8d, 8w	6m	Yes	Yes
Manthey et al.	2016	300	D: STADI	.91	Change	To-do lists	Gratitude list	Intra	84	34	NC (college)			8d, 8w	1m	Yes	No
Mongrain & Anselmo-Matthews	2012	190	D: CES-D	.88	Change	Early memories	Gratitude list	Intra	83	33	NC (community)	Yes	19.53	7d, 1w	1m	Yes	Yes
O'Leary & Dockray	2015	22	D: Edinburgh Depression Scale	.87	Change	WLC	Gratitude list and meditation	Intra	100	28		Yes		12d, 3w		Yes	No
Ozimkowski (Unpublished)	2007	89	D: CES-D for children	.86	Post-test	Yesterday's events	Gratitude visit	Inter	48	13[d]	NC (students in grades 3, 8, & 12)			1d, 1w		No	No
Proyer et al.	2014	108	D: CES-D	.94	Change	Early memories	Gratitude visit and list		100	56	NC (community)	No	14.63	1w	1m	Yes	No
Senf & Liau	2012	76	D: CES-D	.88	Change	WLC	Gratitude visit and list	Comb	70	20	NC (college)	Yes	16.98	7d, 1w	2w	No	Yes

Author	Year	N	Outcome(s)	Reliability (Alpha)[a]	Change Score or Post-test Only	Control	Intvn.	Intvn. Type	% Female[b]	Mean Age[c]	Pop.	Dep. Threshold	Baseline CES-D	Duration (Days, Weeks)	Follow-up	Online	Compliance
Sergeant & Mongrain	2011	514	D: CES-D	.88	Change	Early memories	Gratitude list	Intra	81	34	NC (community)	Yes	21.40	7d, 1w	1m	Yes	Yes
Smullen (Unpublished)	2012	35	D: Geriatric Depression Scale		Post-test	Yesterday's events	Gratitude journal	Intra	80	69	NC (community)			12d, 6w		No	No
Southwell & Gould	2017	109	A, D: DASS-42	.84, .91	Change	WLC	Gratitude list	Intra	88	34	C (current diagnosis of anxiety or depressive disorder)	Yes		9d, 3w		Yes	Yes
Timmons & Ekas	2018	64	D: CES-D	.90	Change	Previous week's activities	Gratitude letter	Inter	100	40	NC (mothers of children with ASD)	Yes	20.80	8d, 8w	1m	Yes	No
Toepfer et al.	2012	183	D: CES-D10	.88	Change	WLC	Gratitude Visit	Inter	86	26	NC (college)	Yes		3d, 3w		No	No
Watkins et al.	2008	86	A: IES-R	.95	Change	Tomorrow's Plans	Gratitude journal for unpleasant memory	Intra			NC (college)			3d, 1w	1w	Yes	No

Author	Year	N	Outcome(s)	Reliability (Alpha)[a]	Change Score or Post-test Only	Control	Intvn.	Intvn. Type	% Female[b]	Mean Age[c]	Pop.	Dep. Threshold	Baseline CES-D	Duration (Days, Weeks)	Follow-up	Online	Compliance
Watkins et al.	2015	87	D: CES-D	.88	Change	Daily semantic memories	Gratitude list	Intra	71		NC (college)	No	14.78	7d, 1w	5w	No	No
Wolfe & Patterson	2017	80	D: CES-D	.91	Change	WLC	Gratitude list	Intra	100	20	NC (college)			14d, 2w		Yes	Yes

Note. Blank cells indicate data was either not reported in the study or it is not-applicable; Unpublished = unpublished thesis or dissertation; N = combined sample size for experimental and control groups; D = Depression; A = Anxiety; WLC = waitlist control; Intvn. = intervention; Intra = intrapersonal; Inter = interpersonal; Comb = combined inter & intrapersonal; Pop. = type of study population; NC = non-clinical; C = clinical; Dep. Threshold = whether combined sample met recommended threshold for clinically significant depressive symptoms; Baseline CES-D = average baseline depressive symptoms for combined sample, 20-item measure only; d = days; w = weeks; m = months; Compliance = whether the authors reported a compliance or adherence check; DASS-21 = Depression Anxiety Stress Scales; CES-D = Center for Epidemiological Studies Depression Scale; CES-D10 = 10 item version of the Center for Epidemiological Studies Depression Scale; PSWQ = Penn State Worry Questionnaire; PHQ-9 = Patient Health Questionnaire-9; IES-R = Impact of Event Scale Revised; HADS = Hospital Anxiety and Depression Scale; STADI = State-Trait-Anxiety-Depression Inventory.

[a] If the authors reported both an anxiety and a depression reliability for a scale (e.g., the DASS-21), the anxiety reliability is listed first.
[b,c] Values are rounded to the nearest whole integer.
[d] Study did not report overall mean age, so mean age was estimated based on sample sizes and typical ages for students in grades 3, 8, and 12 (9, 14, and 18 years old, respectively).
[e] Study had multiple intervention groups with different durations. Therefore, study duration was not included.
[f] The HADS assesses depression and anxiety simultaneously, therefore this study was only included in the overall meta-analysis.

Table 2

Weights and Effect Sizes of Included Studies for the Overall Meta-Analysis at Post-Test

Study Author(s)	**Weight**	**Hedges' *g***
Booker & Dunsmore	3.80%	−0.45
Cheng, Tsui, & Lam	3.09%	−0.64
Gander et al.	4.59%	−0.12
Gavian	4.65%	−0.09
Geraghty Dissertation Study 4	4.25%	−0.31
Geraghty, Wood, & Hyland	3.41%	−1.64
Harbaugh & Vasey	3.84%	0.07
Jackowska et al.	3.41%	−0.35
Kerr, O'Donovan, & Pepping	2.27%	−0.34
Ki	4.20%	−0.96
Lambert, Fincham, & Stillman Study 5	3.87%	−0.36
Lambert, Fincham, & Stillman Study 7	3.35%	−0.16
Lyubomirsky et al.	4.65%	−0.02
Manthey, Vehreschild, & Renner	5.00%	−0.22
Mongrain & Anselmo-Matthews	4.49%	−0.21
O'Leary & Dockray	1.43%	−0.28
Ozimkowski	3.56%	−0.06
Proyer et al.	3.62%	−0.12
Senf & Liau	3.32%	−0.19
Sergeant & Mongrain	5.34%	−0.05
Smullen	2.19%	0.02
Southwell & Gould	4.07%	−0.20
Timmons & Ekas	2.86%	−0.33
Toepfer, Cichy, & Peters	4.44%	−0.30
Watkins et al.	3.47%	−0.04
Watkins, Uhder, & Pichinevskiy	3.53%	−0.05
Wolfe & Patterson	3.32%	−0.60

Table 3

Main Effects of Meta-Analysis

Outcome	K	N	Hedges' *g* [95% CI]	*SE* (*p*)	Heterogeneity Test
Post-test					
Depression	24	3,400	−0.23 [−0.33, −0.13]	0.05 ($p < .01$)	$\tau^2 = 0.02$, Q (23) = 39.36, $p = .02$
Anxiety	5	508	−0.52 [−1.11, 0.08]	0.30 ($p = .09$)	$\tau^2 = 0.41$, Q (4) = 40.20, $p < .01$
Overall	27	3,675	−0.29 [−0.41, −0.16]	0.06 ($p < .01$)	$\tau^2 = 0.07$, Q (26) = 80.05, $p < .01$
Post-test (outliers excluded)[a]					
Depression	23	3,239	−0.17 [−0.24, −0.10]	0.04 ($p < .01$)	$\tau^2 = 0.00$, Q (22) = 18.27, $p = .69$
Anxiety	4	400	−0.16 [−0.38, 0.05]	0.11 ($p = .13$)	$\tau^2 = 0.01$, Q (3) = 3.37, $p = .34$
Overall	25	3,406	−0.18 [−0.25, −0.11]	0.04 ($p < .01$)	$\tau^2 = 0.00$, Q (24) = 19.49, $p = .73$
Follow-up					
Depression	12	2,232	−0.24 [−0.35, −0.12]	0.06 ($p < .01$)	$\tau^2 = 0.01$, Q (11) = 17.19, $p = .10$
Overall	13	2,318	−0.23 [−0.34, −0.12]	0.06 ($p < .01$)	$\tau^2 = 0.01$, Q (12) = 18.33, $p = .11$
Follow-up (outlier excluded)[b]					
Depression	11	2,164	−0.20 [−0.29, −0.11]	0.05 ($p < .01$)	$\tau^2 = 0.00$, Q (10) = 10.67, $p = .38$
Overall	12	2,250	−0.19 [−0.28, −0.10]	0.05 ($p < .01$)	$\tau^2 = 0.00$, Q (11) = 11.62, $p = .39$

Note. K = number of studies included in analysis. An independent follow-up analysis was not conducted for anxiety, as there were only 2 studies assessing anxiety with follow-up data.

[a] Outliers were Geraghty, Wood, & Hyland (2010) for anxiety and Ki (2009) for depression.

[b] Outlier was Cheng, Tsui, & Lam (2015) for depression.

Table 4

Meta-Regression Analyses

Timepoint	Moderator	K (with subgroups) & Total N	Test of Moderation	Subgroup Hedges' *g* (*SE*), *p*
Post-test	Control group	27 (N = 3,675) Active = 18 WLC = 9	$b = -0.32$, $SE = 0.13$, $p = .01$	†Active $g = -0.18$ (0.06), $p < .01$ WLC $g = -0.51$ (0.15), $p < .01$
	Type of intervention	25 (N = 3,232) Interpersonal = 4 Intrapersonal = 19 Combined = 2	Q (2) = 0.80, $p = .67$	†Interpersonal $g = -0.15$ (0.09), $p = .08$ Intrapersonal $g = -0.34$ (0.09), $p < .01$ Combined $g = -0.27$ (0.16), $p = .09$
	Online implementation	27 (N = 3,675) Online = 20 Offline = 7	$b = 0.13$, $SE = 0.15$, $p = .41$	†Online $g = -0.32$ (0.08), $p < .01$ Offline $g = -0.20$ (0.08), $p = .02$
	Publication status	27 (N = 3,675) Published = 22 Not published = 5	$b = -0.03$, $SE = 0.16$, $p = .85$	†Published $g = -0.28$ (0.07), $p < .01$ Not published $g = -0.30$ (0.19), $p = .11$
	Depressive symptoms threshold	18 (N = 2,695) Met = 13 Not met = 5	$b = -0.11$, $SE = 0.14$, $p = .46$	Met $g = -0.26$ (0.08), $p < .01$ †Not met $g = -0.14$ (0.10), $p = .16$
	Baseline CES-D	10 (N = 1,798)	$b = 0.00$, $SE = 0.01$, $p = .74$	
	Year of publication	27 (N = 3,675)	$b = 0.00$, $SE = 0.02$, $p = .86$	
	Percentage female	26 (N = 3,589)	$b = 0.00$, $SE = 0.01$, $p = .95$	
	Mean age	20 (N = 2,817)	$b = 0.00$, $SE = 0.01$, $p = .96$	

Timepoint	Moderator	K (with subgroups) & Total N	Test of Moderation	Subgroup Hedges' *g* (*SE*), *p*
	Duration (days)	25 (N = 3,232)	$b = -0.01$, $SE = 0.01$, $p = .50$	
	Duration (weeks)	26 (N = 3,563)	$b = 0.00$, $SE = 0.03$, $p = .96$	
	Compliance check	27 (N = 3,675) No = 13 Yes = 14	$b = 0.18$, $SE = 0.12$, $p = .15$	†No $g = -0.38$ (0.13), $p < .01$ Yes $g = -0.18$ (0.05), $p < .01$
	Risk of bias (summary score)	27 (N = 3,675)	$b = -0.02$, $SE = 0.05$, $p = .71$	
	Risk of bias (categorical risk)	27 (N = 3,675) Low = 7 Medium = 13 High = 7	Q (2) = 0.83, $p = .66$	†Low $g = -0.23$ (0.07), $p < .01$ Medium $g = -0.35$ (0.14), $p < .01$ High $g = -0.20$ (0.06), $p < .01$
Follow-up	Control group	13 (N = 2,318) Active = 10 WLC = 3	$b = -0.48$, $SE = .14$, $p < .01$	†Active $g = -0.16$ (0.05), $p < .01$ WLC $g = -0.63$ (0.13), $p < .01$
	Type of intervention	11 (N = 1,875) Interpersonal = 2 Intrapersonal = 8 Combined = 1	Q (2) = 1.18, $p = .55$	†Interpersonal $g = -0.25$ (0.12), $p = .04$ Intrapersonal $g = -0.24$ (0.08), $p < .01$ Combined = N/A
	Online implementation	13 (N = 2,318) Online = 11 Offline = 2	$b = -0.10$, $SE = .19$, $p = .62$	†Online $g = -0.22$ (0.06), $p < .01$ Offline $g = -0.32$ (0.22), $p = .14$
	Publication status	13 (N = 2,318) Published = 12 Not published = 1	N/A	N/A

Timepoint	Moderator	K (with subgroups) & Total N	Test of Moderation	Subgroup Hedges' *g* (*SE*), *p*
	Depressive symptoms threshold	11 (N = 1,931) Met = 7 Not met = 4	$b = 0.10$, $SE = 0.15$, $p = .49$	Met $g = -0.21$ (0.07), $p < .01$ †Not met $g = -0.34$ (0.15), $p = .03$
	Baseline CES-D	9 (N = 1,603)	$b = -0.01$, $SE = 0.01$, $p = .40$	
	Year of publication	13 (N = 2,318)	$b = -0.04$, $SE = 0.02$, $p = .06$	
	Percentage female	12 (N = 2,232)	$b = 0.01$, $SE = 0.01$, $p = .07$	
	Mean age	9 (N = 1,903)	$b = 0.01$, $SE = 0.01$, $p = .08$	
	Duration (days)	11 (N = 1,875)	$b = 0.00$, $SE = 0.05$, $p = .94$	
	Duration (weeks)	12 (N = 1,983)	$b = -0.01$, $SE = 0.02$, $p = .70$	
	Compliance check	13 (N = 2,318) No = 4 Yes = 9	$b = -0.10$, $SE = 0.13$, $p = .44$	†No $g = -0.17$ (0.09), $p = .05$ Yes $g = -0.27$ (0.08), $p < .01$
	Risk of bias (summary score)	13 (N = 2,318)	$b = -0.01$, $SE = 0.05$, $p = .87$	
	Risk of bias (categorical risk)	13 (N = 2,318) Low = 4 Medium = 4 High = 5	Q (2) = 0.21, $p = .90$	†Low $g = -0.26$ (0.16), $p = .11$ Medium $g = -0.23$ (0.07), $p < .01$ High $g = -0.21$ (0.09), $p = .02$

Note. All analyses were performed using the overall aggregated effect size, except for depressive symptoms threshold and baseline CES-D, for which the depression-specific effect size was used. N/A = test was not applicable. K = number of studies included in analysis, with subgroup if applicable; *b* = unstandardized regression coefficient for meta-regression; WLC = waitlist control; Depression threshold = whether combined sample met recommended threshold for clinically significant depressive symptoms; Baseline CES-D = average baseline depressive symptoms for combined sample, 20-item measure only; † = reference group for categorical moderator in meta-regression. Subgroup Hedges' *g* is listed only for categorical moderators.

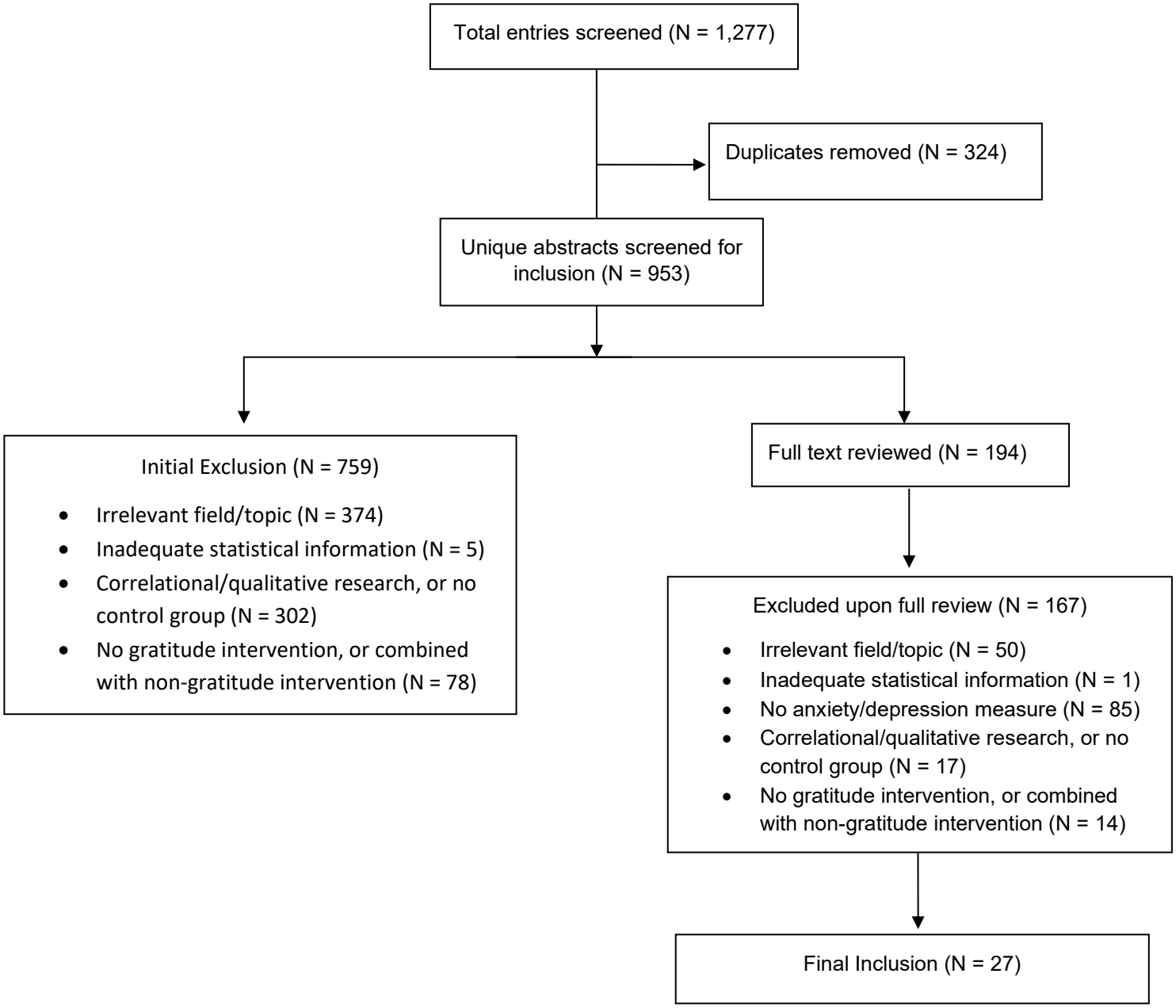

Fig 1 Flowchart of the study inclusion process. Template adapted from "The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials," by D. Moher, K. Schulz, and D. Altman, 2001, *Lancet*, 357, p. 1193. Journals publishing the original CONSORT flowchart have waived copyright protection

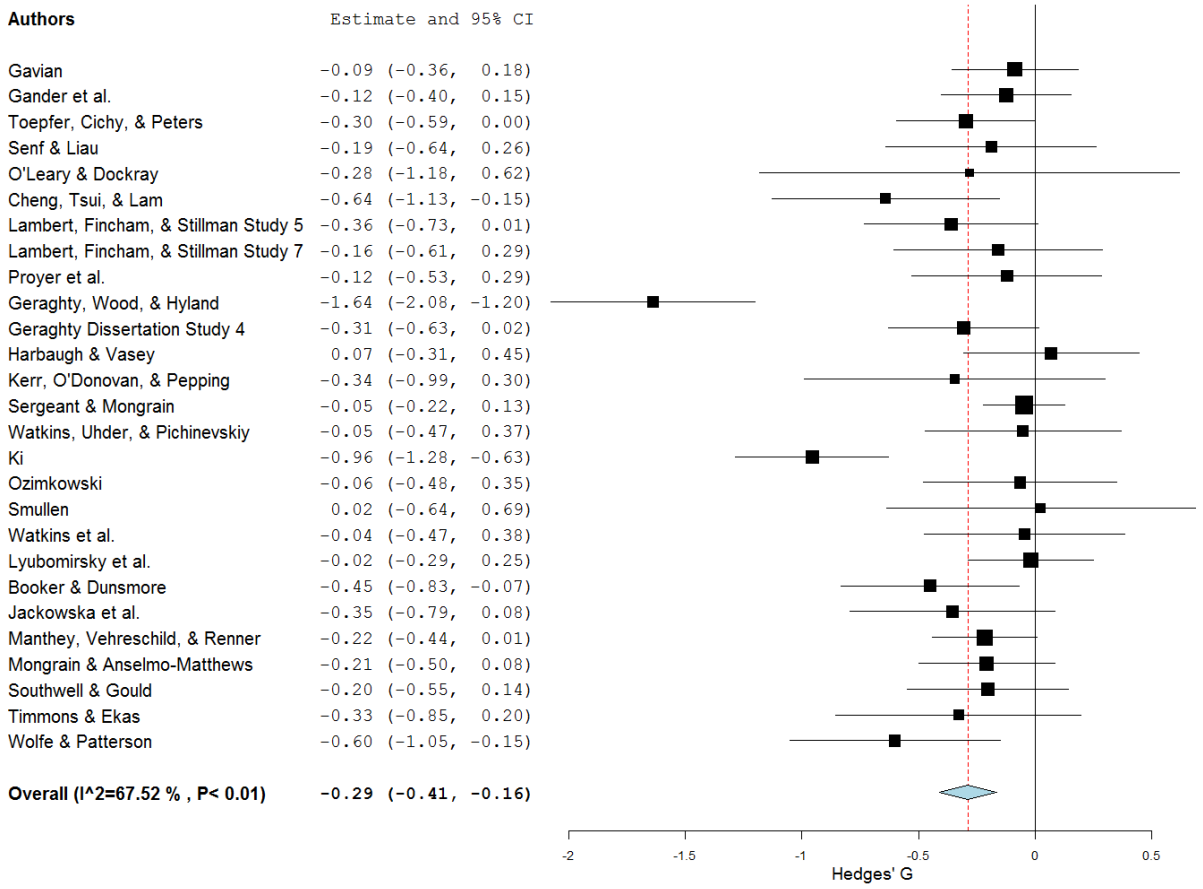

Fig 2 Forest plot of included studies for overall post-test analysis. Squares are individual effect sizes with their corresponding 95% CI indicated by the horizontal lines. The diamond is the overall 95% CI for all studies

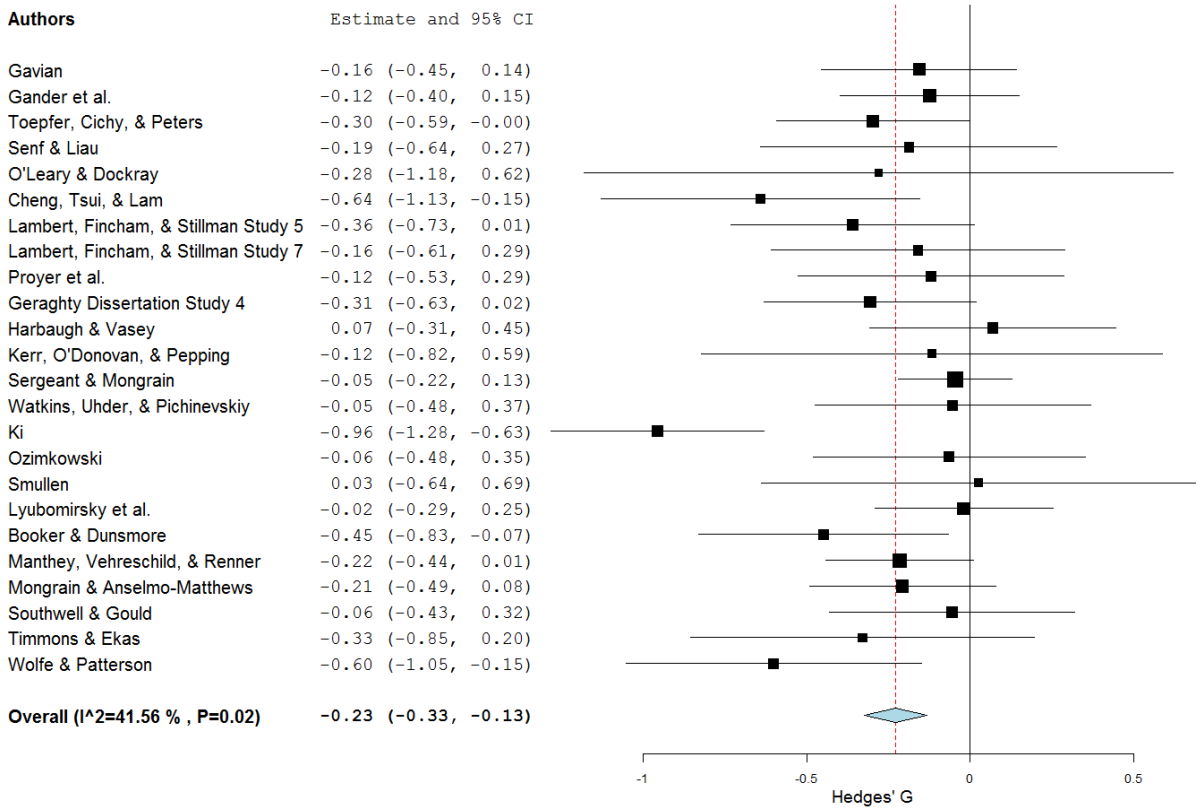

Fig 3 Forest plot of included studies for depression post-test analysis. Squares are individual effect sizes with their corresponding 95% CI indicated by the horizontal lines. The diamond is the overall 95% CI for all studies

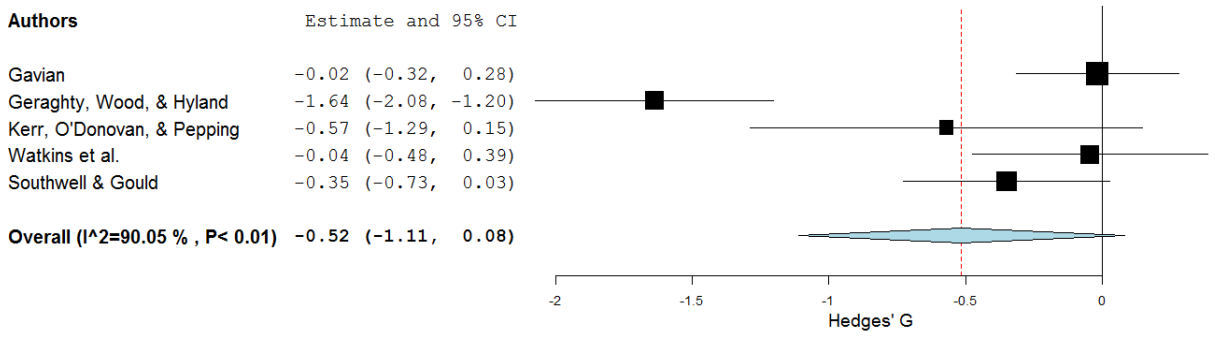

Fig 4 Forest plot of included studies for anxiety post-test analysis. Squares are individual effect sizes with their corresponding 95% CI indicated by the horizontal lines. The diamond is the overall 95% CI for all studies

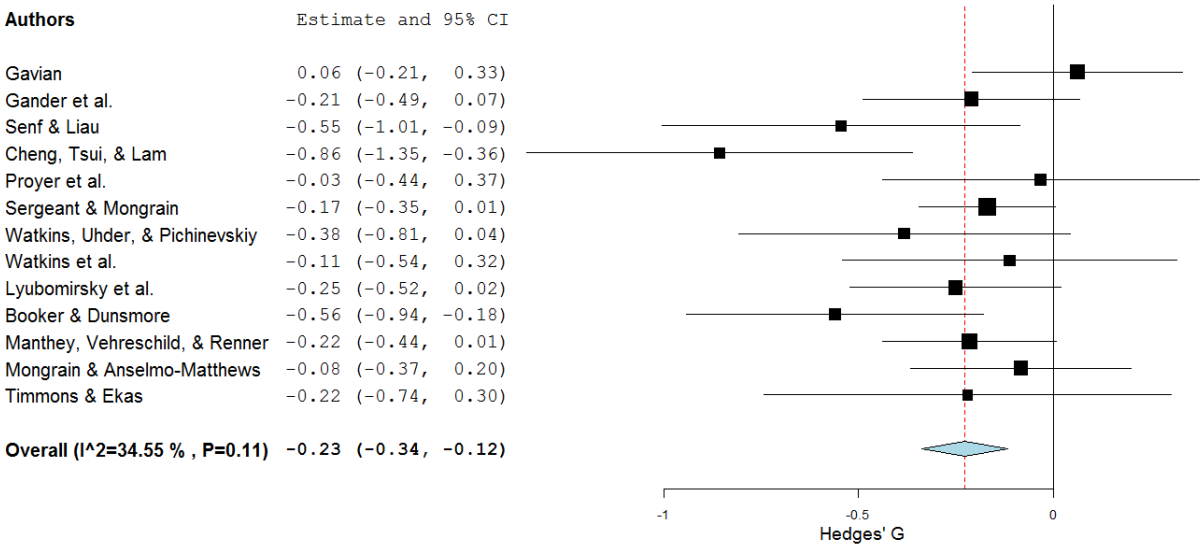

Fig 5 Forest plot of included studies for overall follow-up analysis. Squares are individual effect sizes with their corresponding 95% CI indicated by the horizontal lines. The diamond is the overall 95% CI for all studies