Feeling Thanks and Saying Thanks: A Randomized Controlled Trial Examining If and How Socially Oriented Gratitude Journals Work

Brenda H. O'Connell,[1] Deirdre O'Shea,[2] and Stephen Gallagher[1,3]

[1] Centre for Social Issues Research, Study of Anxiety, Stress and Health Lab, University of Limerick
[2] Kemmy Business School, University of Limerick
[3] Health Research Institute, University of Limerick

Objective: This study examined the effect of a reflective interpersonal gratitude journal, a reflective-behavioral interpersonal gratitude journal and an active control journal, on primary qualities of well-being and depression. **Method:** Participants ($n = 192$; 67.2% female) completed this 3-month longitudinal randomized controlled design. **Results:** Participants in the reflective-behavioral condition experienced the greatest improvements in affect balance and reductions in depression at immediate posttest. Both gratitude interventions improved affect balance at 1 month, compared to the control. Changes in affect balance for those in the reflective-behavioral condition were mediated by the rate at which people expressed gratitude in their existing relationships. This effect was moderated by participant's baseline depressive status. **Conclusion:** Expressing felt gratitude to others appears to be a crucial step in deriving benefits, and these benefits may not be limited to the emotionally healthy. Given the applied popularity of gratitude interventions, understanding not only *if* but also *how* they work is essential. © 2017 Wiley Periodicals, Inc. J. Clin. Psychol. 73:1280–1300, 2017.

Keywords: gratitude; psychological intervention; randomized controlled trial; well-being

Recently, there has been increased effort focused on studying psychological variables that support health and well-being, rather than solely focusing on factors that contribute to mental illness and disease (Seligman & Csikszentmihalyi, 2014). As a result of this and the increasing evidence that the benefits of subjective well-being greatly exceeds feeling good in and of itself (Lyubomirsky, King, & Diener, 2005), the past decade has witnessed growing research interest in positive psychological interventions (PPIs) as a way of improving key markers of mental health. At the forefront of PPIs and regarded as the most established are gratitude interventions (Parks & Schueller, 2014). Underpinning PPIs is a proposed positive activity model that outlines the conditions under which interventions are most efficacious (Lyubomirsky & Layous, 2013). This model reveals caveats in empirical evidence regarding the role of activity features (e.g., reflective or behavioral) and characteristics of the person engaging in the activity (e.g., baseline affective state); key factors identified as influencing intervention outcomes. Therefore, examining both mediational and moderating factors, in addition to outcomes of interventions is essential in order to gain a deeper understanding of *how* and *for whom* PPIs work best. This information is particularly needed for gratitude interventions, given their unprecedented applied popularity, to assist in their optimal application.

Gratitude Interventions

Gratitude interventions aim to promote a sense of gratefulness and are often characterised as an "other-oriented" way of enhancing well-being (Davis et al., 2016). In 2003, Emmons

Please address correspondence to: Brenda H. O'Connell, Centre for Social Issues Research, Study of Anxiety, Stress and Health Lab, Department of Psychology, University of Limerick, Ireland. E-mail: brenda.oconnell@ul.ie

and McCullough (2003) conducted the first set of studies demonstrating the effectiveness of gratitude journals. Subsequent evaluations of gratitude journals and behavioral displays of gratitude have shown similar increments in life satisfaction and decreases in negative affect compared to controls (Cheng, Tsui, & Lam, 2015; Kerr, O'Donovan, & Pepping, 2015; Krejtz, Nezlek, Michnicka, Holas, & Rusanowska, 2016; Redwine et al., 2016; Seligman, Steen, Park, & Peterson, 2005). However, despite their acceptance, the research underlying this evidence base has recently faced methodological critique due to poor experimental trial quality and insufficient control groups (Mohr et al., 2009; Parks, 2014) and reporting (Bolier et al., 2013). Recent meta-analytical findings (Davis et al., 2016), albeit on their short-term efficacy, highlight the need to temper enthusiasm until longer, more rigorous interventions demonstrate stronger evidence of efficacy. Thus, strict adherence to preregistered trial protocol and adherence to the Consolidated Standards of Reporting Trials (CONSORT; Schulz, Altman, & Moher, 2010) were adopted to ensure accuracy, transparency, and rigor in the current investigation.

Gratitude and Relationships

Gratitude is defined as a positive, social emotion (Emmons & Mishra, 2011) that is robustly associated with features of high-quality, healthy social relationships (Algoe, Haidt, & Gable, 2008; O'Connell, O'Shea, & Gallagher, 2016a,b; Wood, Maltby, Gillett, Linley, & Joseph, 2008) and fosters socially productive behaviors (Wood, Froh, & Geraghty, 2010). Although there are many diverse aspects to gratitude, in many instances it can be depicted as other oriented in nature (Davis et al., 2016). This inherently social/relational aspect of gratitude may predict unique pathways to well-being (Emmons & Mishra, 2011) through the building of social connections (Fredrickson, 2004). Indeed, recent cross sectional and experimental evidence has demonstrated the indirect effect of gratitude on life satisfaction through social support and friendship quality, respectively (Kong, Ding, & Zhao, 2015; O'Connell, O'Shea, & Gallagher, 2016c). These findings indicate that cultivating gratitude and directing it toward other people may be integral in the promotion of satisfying relationships, and consequently health and well-being, and as such will be the focus of the current investigation.

Features of the Gratitude Intervention

According to Emmons and Mishra (2011), gratitude is experienced when there is "an acknowledgement that we have received something of value from others" (p. 248), and although gratitude is often depicted as other oriented, in many cases it is never expressed to the other. Limited research (for an exception, see Davis et al., 2016) has examined whether or not including a behavioral component (i.e., expressing gratitude to another) in a gratitude intervention causes unique and stronger effects than only a cognitive component (i.e., grateful reflection and contemplation), despite this being highlighted as a key activity characteristic that may influence PPI efficacy, and gratitude interventions in particular (Lyubomirsky & Layous, 2013).

Studies have shown that both behavioral and reflective activities can be effective (Boehm, Lyubomirsky, & Sheldon, 2011; Emmons & McCullough, 2003; Froh, Kashdan, Ozimkowski, & Miller, 2009; Rash, Matsuba, & Prkachin, 2011), but have not contrasted these. In an exception to this, Davis et al. (2016) examined whether the type of gratitude intervention (i.e., gratitude journals or grateful expression) moderated the relationship between the gratitude intervention and effect size on psychological well-being. It was predicted that expressions of gratitude would outperform gratitude journals/lists due to the requirement for interpersonal and social interaction. This was not supported as both types of gratitude interventions showed similar efficacy. Nonetheless, experimental manipulations of gratitude within the area of romantic relationships clearly evidences the value of expressing gratitude to others. It has been shown to increase positive perceptions and communal strength of the romantic relationship, improve aspects of relationship maintenance, and improve marital satisfaction compared to control conditions (Gordon, Arnette, & Smith, 2010; Lambert, Clark, Durtschi, Finham, & Graham, 2010, Study 3; Lambert & Fincham, 2011, Study 4; Schramm, Marshall, Harris, & Lee, 2005).

Therefore, Davis et al. (2016) still contends that individuals may not experience optimal benefits of gratitude interventions unless there is a shared component. Thus far, there is no evidence for whether or not a behavioral component in addition to a reflective component may be more effective than just reflection in isolation. It is possible that a gratitude intervention that aims to cultivate both an inward attitude of gratitude, and subsequently an outward expression of that gratitude, may lead to optimal benefits. The current research examines whether a gratitude journal that cultivates interpersonal, other-oriented gratitude is as effective in enhancing well-being, as when this gratitude is subsequently expressed to the other; therefore, containing both a reflective and behavioral component (see Hypothesis 1).

Additionally, the level to which people express gratitude in their existing relationships, as opposed to overall levels of gratitude, is an important proximal mediator of any desired outcomes, and thus was examined in this research (see Hypothesis 2). Examining activity features such as these is an important endeavour as they influence the interventions success at improving desirable outcomes and provide validation information for the positive activity model (Lyubomirsky & Layous, 2013), contributing to optimal application.

Features of the Person

In addition to examining activity features, studies are beginning to show that features of the person are also significant in determining intervention efficacy (Briner & Walshe, 2015; Clauss et al., in press; Proyer, Wellenzohn, Gander, & Ruch, 2015). For example, personality traits, such as extraversion and openness have shown to enhance the impact of PPI effects on happiness and depression (Senf & Liau, 2013) and preexisting levels of social support were not found to impede efficacy of a PPI targeting relationships (O'Connell et al., 2016a). Of particular importance to the potential effectiveness of these interventions are people's initial affective states. Research examining a representative online sample of "happiness seekers" found that up to half of the people intentionally trying to become happier had levels of distress equal to that of clinical depression (Parks, Della Porta, Pierce, Zilca, & Lyubomirsky, 2012), indicating that many people with depressive symptoms are already engaging in self-administered PPIs. Consequently, there is a need to validate the extent to which specific interventions may be efficacious for these individuals.

One of the first studies showing the efficacy of brief online PPIs, conducted by Seligman and colleagues (2005), comprised 411 volunteers who were mildly depressed on average. Participants who either wrote about three good things during the day or used their signature strengths in a new way, over a 1-week period, experienced lasting improvements in happiness and depressive symptomology for up to 6 months later. In contrast to this, another study examining the benefits of PPIs for individuals who were mildly depressed or "dysphoric" (Sin, Della Porta, & Lyubomirsky, 2011) found that the practice of delivering a gratitude letter actually caused reductions in well-being at immediate posttest. The authors proposed that the process of writing a gratitude letter to someone that they had never thanked properly may have been viewed as a frustrating and difficult task for these individuals. In their meta-analysis of 51 PPIs, Sin and Lyubomirsky (2009) found that depression status moderated the effectiveness of the PPIs, such that those who had elevated scores in depression experienced greater improvements in well-being and reductions in depressive symptoms compared to those who were not.

Although not consistent–these findings challenge the predominant idea that people with depressive symptoms might not be suitable for PPIs because of their characteristic cognitive, affective, and behavioral challenges (Sin & Lyubomirsky, 2009)–they may in fact be particularly useful as self-help for people scoring on the milder end of depressive symptomology scales. Given these varied findings, there are calls for more research identifying specific PPIs that are suited to individuals within specific affective ranges (Lyubomirsky & Layous, 2013; Senf & Liau, 2013). As with any intervention, there may be subgroups for which there may be distinct causal pathways (Hinshaw, 2002). Given the potential for gratitude interventions to directly foster protective factors and mitigate proximal risk factors for psychological disorders (Layous, Chancellor, Lyubomirsky, Wang, & Doraiswamy, 2011; Layous, Chancellor, & Lyubomirsky, 2014; O'Connell et al., 2016b), more research is needed to examine their effects in a nonclinical

sample within specific affective ranges. Therefore, this research examined whether participant's initial depression scores moderated the effects found (see Hypothesis 3).

The Current Study

Using a longitudinal randomized controlled design, the present study examined the effect of (a) a reflective interpersonal gratitude journal designed to foster gratitude for other people; (b) a reflective-behavioral interpersonal gratitude journal designed to foster gratitude for other people and subsequently expression of that gratitude; and (c) an active control journal–on gratitude, gratitude expression, relationship quality, and primary qualities of well-being–affect balance, life satisfaction, and depression (in keeping with previous studies; Davis et al. 2016).

The following hypotheses were generated:

H1: Participants who completed the reflective-behavioral gratitude journal would experience greater improvements in outcomes (affect balance, life satisfaction, depression, gratitude, gratitude expression, relationship satisfaction) at immediate posttest, 1-month and 3-month follow-up compared to those who completed the comparable reflective-only gratitude journal and control journal.

H2: Outcomes for the reflective-behavioral gratitude journal would be mediated by the level to which people expressed gratitude in their existing relationships.

H3: These mediated effects will differ depending on participant's baseline depressive status.

Method

Design

This was a double-blind, randomized controlled group study with an approximate 1:1:1 allocation ratio, following the CONSORT guidelines (Schulz et al., 2010). Participants were randomized using a random sequence generator and allocation was concealed by placing the experimental journals or the control journal in sequentially numbered identical opaque sealed envelopes, organized by an independent research assistant. In this way, both the participants and the investigators enrolling participants were blind to group allocation across all assessments. Group allocation was revealed to the principal investigator following the final assessment. Life satisfaction served as the primary outcome.

Participants

A convenience sample of 192 participants, 67.2% female, were recruited. At the randomization stage, this comprised students (70.8%) and nonstudents (28.6%), with one person not identifying their student status. In terms of nationality, the majority of participants identified as Irish (90.6%), 3.1% European, 2.6% American, 2.6% Asian, 0.5% Canadian, and one missing case. Nearly half of the participants were single (47.4 %), while 38% reported that they were in a relationship (unmarried), 11.5% married, 0.5% separated or divorced, and the remaining 2.6% did not report their relationship status. The sample were aged between 18 and 84 years (mean [M] = 27.08, standard deviation [SD] = 12.63), with the majority aged between 18 and 30 years (78.6%). Inclusion criteria on enrolment were people aged 18 years or older (to exclude children) and English language reading and writing proficiency (all the questionnaires employed were designed for English speakers). Ethical approval for this study was granted by the host university's research ethics committee. All those recruited provided written informed consent prior to participation. All those who completed each part of the study were entered into a draw for €50.

Procedure

Participants were recruited via snowball sampling (nonstudent participants) or via an e-mail sent to all students in the host university, inviting them to take part in the study named Writing

Journals and Wellbeing. Participants came to a lab located on the university campus and were given information sheets outlining the requirements of participating in the research and that the purpose of the research was to better understand journal writing and well-being. Participants provided written consent before commencing the study and completing the questionnaires. Before this, an independent research assistant randomized 200 questionnaire packs and journals into identical opaque sealed labelled envelopes, to achieve double blind design and conceal allocation sequence from the investigators in advance. As all 200 packs were not distributed, this led to a slightly uneven distribution between the conditions.

Participants completed baseline questionnaires and were randomly assigned to either the reflective-behavioral, reflective-only, or the control condition. Distinct journals were designed for each condition, adapted from previous research (O'Connell et al., 2016c) and distributed with the questionnaire pack. These journals contained intervention guidelines and separate pages for each of the 9 days participants were required to write.

In the reflective-behavioral and reflective-only condition, participants were given the following instructions on the first page of their journal:

> Writing is a great way to reflect on your daily events, helping to look back and focus on the good things in our lives. There are many things in our lives, big and small, which we might be grateful for. For the next three weeks, 3 times a week (9 days in total), reflect back on your day and think of the people you met and interacted with and are grateful for. Please write down in the space provided a number of positive social interactions over the day or friendships/relationships you are grateful for.

The reflective-behavioral condition contained the following additional statement:

> At the end of each week, express this gratitude to a person of your choice face-to-face or through e-mail, facebook, a kind note, tell him/her how much you appreciate something specific that he/she does and reflect on their reaction and how you feel.

In the control condition, participants were given the following instructions on the first page of their journal:

> Writing is a great way to reflect on your daily events, helping us to look back and reflect on our lives. For the next three weeks, 3 times a week (9 days in total), please write in the spaces provided things that happened during the day.

This control intervention was designed so that it would likely generate an equivalent degree of expectancy as the experimental interventions. Therefore, the same introduction message "Writing is a great way to reflect on your daily events, helping us to look back and reflect on our lives" was retained to isolate the theorized active ingredient of the experimental interventions, in this case interpersonal gratitude. As noted by Wood and colleagues (2010), the most suitable comparison groups are ones that are identical to the experimental groups in all aspects apart from the feature of interest (interpersonal gratitude), to discount other generic explanations. Control groups of a similar nature (reflecting and writing about typical things that occur during a day) have been employed in previous research (Froh et al., 2009; Sheldon & Lyubomirsky, 2006) and do not appear to significantly influence well-being beyond psychological expectancy of change.

Participants were given a choice of what days of the week they wanted to complete their journal, to foster an autonomy supportive environment (Della Porta et al., 2012 as cited in Lyubomirsky & Layous, 2013). They were instructed to read their journal instructions carefully and informed that they would receive text messages on three occasions reminding them to complete their journal entries, to improve adherence. After the 3-week intervention period, participants completed assessments, as previously administered at baseline, at immediate posttest, and 1 month and 3 months later. In the final 3-month follow-up, participants who dropped out of the study were asked to please indicate why by ticking the most appropriate option from a list of reasons. Upon completion, participants were thanked and debriefed.

Measures

Participants completed demographics that included age, gender, nationality, relationship status, and student status. Self-report measures were completed in paper-and-pencil format at baseline and either paper-and-pencil or online (to reduce attrition, participants identified their preference) at the three follow-up assessments (immediate posttest at 3 weeks and 1 month and 3 months after intervention cessation). Several questions, guided by the positive activity model (Lyubomirsky & Layous, 2013), were asked to control for person characteristics that may affect the extent to which participants would benefit from the interventions.

To assess the level of expectation of treatment efficacy across both groups of perceiving benefit, participants rated the question "How well do you expect to feel after taking part in the writing activities?" on a 5-point Likert-type scale ranging from 1 (*not at all well*) to 5 *(extremely well)*. At immediate posttest, participants were asked to indicate how many days they completed a diary/journal entry. At follow-up, participants rated the questions "In general, how much do you like writing?" and "How much did you enjoy this writing task?" on a 5-point Likert-type scale ranging from 1 (*not at all*) to 5 *(extremely)*. Participants in the intervention conditions were asked to indicate whether they predominately focused their gratitude on family members, nonfamily members; friends, nonfamily members; boyfriend/girlfriend; or a mixture of all.

The outcomes reported in this paper included assessments of life satisfaction, affect, depression, gratitude, relationship satisfaction, and expression of gratitude in relationships (gratitude expression). All outcomes are outlined in the trial registration protocol (ClinicalTrials.gov NCT02570685).

Gratitude. Gratitude was assessed using the Gratitude Questionnaire-Six Item Form (McCullough, Emmons, & Tsang, 2002). This is a self-report scale measuring the strength and frequency with which participants experience gratitude. Respondents indicated how much they agreed with six statements (Time 1 [T1]: past month; Time 2 [T2]: past 3 weeks; Time 3 [T3]: past month; Time 4 [T4]: past month), for example, "I have so much in life to be thankful for," two of which were reverse scored, for example, "When I look at the world, I don't see much to be grateful for". Respondents rated their answers on a 7-point scale ranging from 1 (*strongly disagree)* 7 to (*strongly agree)*. All items were summed to produce a total gratitude score for each time point. The scale was reported to have high internal consistency in McCullough and colleagues (2002) original study (Cronbach's alpha = .82). Similar alpha coefficients were found in the current study at baseline, posttest, and 1-month and 3-month follow-up, 0.73, 0.78, 0.81, and 0.84, respectively.

Life satisfaction. Life satisfaction was assessed using the Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin, 1985). This is a brief and reliable five-item measure of participants' global perspective of their life satisfaction, which is robustly linked to various mental health and clinical outcomes (Lyubomirsky et al., 2005) and employed across different disciplines (Pavot & Diener, 2008). This scale is also as sensitive to occasion specificity as much longer scales, such as the Steen Happiness Index (Kaczmarek, Bujacz, & Eid, 2015), and it has the ability to detect change after psychosocial intervention (Pavot & Diener, 1993).

Participants rated their level of agreement on items (T1: past month; T2: past 3 weeks; T3: past month; T4: past month), for example, "In most ways my life is close to my ideal," on a 7-point scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree)*. An overall score was computed by summing the responses of each item. Scores ranged from 5 to 35, in which increments in scores corresponded to increments in satisfaction with life. This scale demonstrates high internal consistency, with alpha coefficients between 0.79 and 0.89 (for review, see Pavot & Diener, 1993). The current study reflected this, with high alpha coefficients at baseline, posttest, and 1-month and 3-month follow-up, 0.85, 0.88, 0.87, and .91, respectively.

The Scale of Positive and Negative Experiences (SPANE; Diener et al., 2010). The SPANE is a 12-item self-report questionnaire that measures positive and negative experiences and affect. Respondents rated how much of the time (T1: past month; T2: past 3 weeks; T3:

past month; T4: past month) they have experienced a combination of general and specific affect (six positive and six negative), for example, "happy," "joyful," on a 5-point frequency scale ranging from 1 (*very rarely or never*) to 5 (*very often or always*). The six positive items and six negative items can be scored separately to create two subscales–positive and negative affective experiences. An overall affect balance score is produced by subtracting summed negative affect scores from summed positive affect scores; the resultant difference score can vary from −24 to 24. Diener and colleagues (2010) reported high internal reliability (Cronbach's alpha = .81− .87) and the current study observed similar levels at baseline, posttest, and 1-month and 3-month follow-up, 0.80, 0.85, 0.86, and .89, respectively, for positive affect, and 0.78, 0.83, 0.87, and 0.87, respectively, for negative affect.

Centre for Epidemiological Studies Short Depression Scale (CES-D 10). This is a 10-item self-report adult symptom survey. The instrument was designed by Radloff (1977) and measures the behavioral, cognitive, and affective symptoms of depression over the previous week, for example, "I was bothered by things that don't usually bother me" and "My sleep was restless," two of which are reverse scored, for example, "I was happy." Participants rated the frequency of the occurrence of each symptom (T1: past month; T2: past 3 weeks; T3: past month; T4: past month) on a 4-point scale ranging from 0 (*rarely or none of the time*) to 3 (*most or all of the time*). Scores ranged from 0 to 30, with higher scores indicating more depressive symptoms. A score of 10 or greater is considered above normative levels and is indicative of "significant" or "mild" depressive symptomatology (Andresen, Malmgren, Carter, & Patrick, 1994). According to these standard cutoffs, 37% (n = 71) of the sample presented with significant depressive symptoms, with 57.3% reporting normative levels and 11 missing cases. Irwin, Artin, and Oxman (1999) reported high internal reliability (Cronbach's alpha = .80) and the current study observed similar levels at baseline, immediate posttest, and 1-month and 3-month follow-up, 0.73, 0.81, 0.86, and 0.86, respectively.

Expression of Gratitude in Relationships (Lambert et al., 2010). Respondents rated three questions (T1: past month; T2: past 3 weeks; T3: past month; T4: past month), for example, "I express my appreciation for the things that my friends/partner does for me" on a 5-point frequency scale ranging from 1 (*never*) to 5 (*very frequently*). Lambert and colleagues (2010) reported high internal reliability (Cronbach's alpha = .94) and the current study observed similar levels at baseline, immediate posttest, and 1-month and 3-month follow-up, 0.82, 0.86, 0.9, and 0.9, respectively. This measure of gratitude expression was used as a potential mediator variable of intervention effects on well-being.

Sample Size Calculation

Using G*Power (version 3.1; Faul, Erdfelder, Lang, & Buchner, 2007; Faul, Erdfelder, Buchner, & Lang, 2009), a priori sample size calculation with a medium partial eta-squared (η^2_p) effect size of 0.06 and power of 0.9 yielded a required sample size of 99, with 33 in each trial arm. Based on previous work (O'Connell et al., 2016a), high attrition rates were predicted; therefore, a sample size of 160 was set as a minimum target. Recruitment ended after one month to ensure follow-ups occurred during the university semester as some participants were recruited on campus.

Data Analytic Strategy

IBM SPSS (version 21) was used to conduct all statistical analysis. Descriptive analyses of baseline homogeneity between the three conditions and attrition bias were examined using univariate analysis of variance and chi-square analysis, to check the randomisation manipulation. Due to the high rates of attrition in PPI studies (Bolier et al., 2013) and in the present study (52.6% at final follow-up), analyses of complete-cases only, using general linear model repeated measures for changes between conditions across time, were conducted separately for each time

point relative to baseline to retain power, with partial eta squared ($\eta 2p$) as a measure of effect size. Where appropriate, post hoc analyses were conducted using pairwise comparisons.

Further, because of the high percentage of missing values due to attrition, a sensitivity analysis was conducted using multilevel modeling, in which an intention-to-treat (ITT) approach was employed with inclusion of all randomised participants. This analysis was run as a sensitivity test on any statistically significant findings obtained in the mixed repeated measures analyses to determine the robustness of such findings. This analysis was completed using the linear mixed model feature in IBM SPSS 21, specifying a linear growth trend, in which maximum likelihood estimations were used to handle missing data.

Simple mediation and moderated mediation analyses using the custom dialogue PROCESS for SPSS (Hayes, 2013) was conducted, to examine if any favorable changes in primary outcomes for the reflective-behavioral gratitude journal would be mediated by the level to which people expressed gratitude in their existing relationships and whether these mediated effects would be moderated by baseline depressive status. Results are reported for bootstrap significance tests using a bias-corrected and accelerated (BCa) 95% confidence interval (CI) with a resample procedure of 5,000 bootstrap samples, whereby an estimate is statistically significant at $p < 0.05$ (two-tailed) or if the 95% CI does not contain zero.

Results

Preliminary Descriptive Analyses

Attrition. Figure 1 displays the consolidated standards of reporting trials (CONSORT) diagram for the trial accrual and retention. Of the 192 participants who consented to participate, 192(100%), 135 (70.3%), 109 (56.8%), and 91 (47.4%) returned at immediate posttest and 1-month and 3-month follow-up, respectively. There was no differential dropout between the control ($n = 31$), reflective-behavioral ($n = 32$), and reflective-only ($n = 38$) conditions, $\chi^2(2) = .454$, $p = 0.797$. Of those who did not complete all parts of the study, three participants indicated that they could not continue because of time restraints.

Participants who did not complete all four time points did not differ from those who completed all four parts ($n = 91$) with respect to demographic characteristics: gender, $\chi^2(1) = 1.411$, $p = 0.235$; age, $F(1, 184) = 0.377$, $p = 0.540$; baseline life satisfaction scores, $F(1, 186) = 0.108$, $p = 0.743$; affect balance, $F(1, 180) = 0.404$, $p = 0.526$; depressive symptoms, $F(1, 179) = 0.00$, $p = 0.989$; gratitude, $F(1, 187) = 0.342$, $p = 0.56$; gratitude expression, $F(1, 188) = 0.00$, $p = 0.989$; or relationship satisfaction, $F(1, 186) = 0.130$, $p = 0.719$.

Tests of baseline homogeneity were conducted and no significant differences between treatment groups were found in sex, $\chi^2(2) = .195$, $p = .907$; age, $F(2, 183) = 1.126$, $p = 0.327$; nationality, $\chi^2(8) = 3.860$, $p = .870$; relationship status, $\chi^2(8) = 5.974$, $p = .650$; student status, $\chi^2(2) = 1.46$, $p = .480$; life satisfaction scores, $F(2, 185) = 0.434$, $p = 0.649$; affect balance, $F(2, 179) = .394$, $p = 0.675$; depressive symptoms, $F(2, 178) = 2.432$, $p = 0.091$; gratitude expression, $F(2, 185) = .679$, $p = 0.508$; or relationship satisfaction, $F(2, 183) = 0.455$, $p = 0.635$. This confirmed successful randomization. Outcome and demographic descriptive characteristics across treatment groups are presented in Table 1.

Intervention compliance and potential confounds. Before beginning the intervention, participants had similar beliefs about treatment efficacy, $F(2, 188) = .164$, $p = .849$, regardless of whether they were in the control ($M = 3.58$, $SD = .83$), reflective-only ($M = 3.63$, $SD = .75$), or reflective-behavioral intervention ($M = 3.56$, $SD = .76$). There were no significant differences between the three conditions on the number of journal entries completed, $\chi 2(18, n = 135) = 17.88$, $p = 0.464$. For the control intervention, the mean number of journal entries was 6.85 days (*median* = 8, $SD = 2.60$); for the reflective-behavioral intervention, the mean number of journal entries was 6.8 days (*median* = 8, $SD = 2.74$); and for the reflective-only intervention, the mean number of journal entries was 7.17 days (*median* = 8, $SD = 2.14$). One participant in the reflective-behavioral condition self-reported that they did not complete any of the journal entries and was therefore excluded from the complete case analysis.

Table 1
Outcome Descriptive Characteristics Across Treatment Groups

	Reflective-behavioural gratitude				Reflective-only gratitude				Control			
	T1 (n=63) M (SD)	T2 (n=40) M (SD)	T3 (n=39) M (SD)	T4 (n=31) M (SD)	T1 (n=68) M (SD)	T2 (n=47) M (SD)	T3 (n=36) M (SD)	T4 (n=30) M (SD)	T1 (n=61) M (SD)	T2 (n=48) M (SD)	T3 (n=34) M (SD)	T4 (n=30) M (SD)
Life satisfaction	24.73 (5.97)	26.32 (5.86)	26.18 (5.33)	24.61 (6.39)	25.52 (5.72)	25.2 (5.77)	25.75 (5.67)	25.00 (6.79)	24.64 (5.8)	25.28 (5.97)	26.47 (4.49)	24.7 (6.86)
Affect balance	8.83 (6.12)	10.53 (7.71)	10.71 (7.64)	9.00 (5.92)	8.92 (5.72)	9.61 (6.28)	11.83 (6.93)	8.9 (6.99)	9.69 (5.74)	8.72 (5.85)	10.06 (5.62)	9.23 (8.05)
Gratitude	35.49 (4.74)	35.89 (4.66)	35.68 (5.52)	35.06 (6.16)	36.33 (4.47)	35.24 (4.70)	36.37 (5.26)	36.1 (5.60)	36.33 (3.97)	35.91 (5.37)	35.68 (5.55)	34.3 (5.58)
Depression	9.88 (4.81)	7.44 (4.96)	8.55 (6.08)	8.16 (5.1)	8.11 (4.09)	7.23 (4.48)	6.22 (4.59)	7.07 (4.91)	8.72 (4.48)	7.79 (5.17)	7.24 (5.31)	9.37 (6.42)
Gratitude Expression	12.29 (3.81)	12.57 (3.43)	12.68 (3.64)	12.29 (4.19)	11.83 (3.55)	12.07 (4.73)	11.86 (3.47)	12.1 (4.09)	12.05 (3.84)	12.15 (3.95)	12.29 (3.9)	12.63 (4.29)
Relationship Satisfaction	19.56 (1.96)	20.38 (1.86)	20.24 (1.54)	19.52 (2.28)	19.88 (2.52)	19.35 (2.15)	19.86 (2.68)	19.57 (2.4)	19.24 (2.07)	18.62 (1.88)	19.5 (1.93)	18.77 (1.69)
Age	26.93 (11.97)				26.0 (9.60)				28.36 (15.77)			
	N (%)				N (%)				N (%)			
Sex (female)	42 (66.7)				47 (69.1)				40 (65.6)			
Relationship Status (single)	31 (49.2)				35 (51.5)				25 (41.0)			
Student	41 (65.1)				50 (73.5)				45 (73.8)			
Nationality (Irish)	56 (88.9)				62 (91.1)				56 (91.8)			

Note. M = mean; SD = standard devoation. *n* denotes the number of participants that completed the respective time point at least in part.
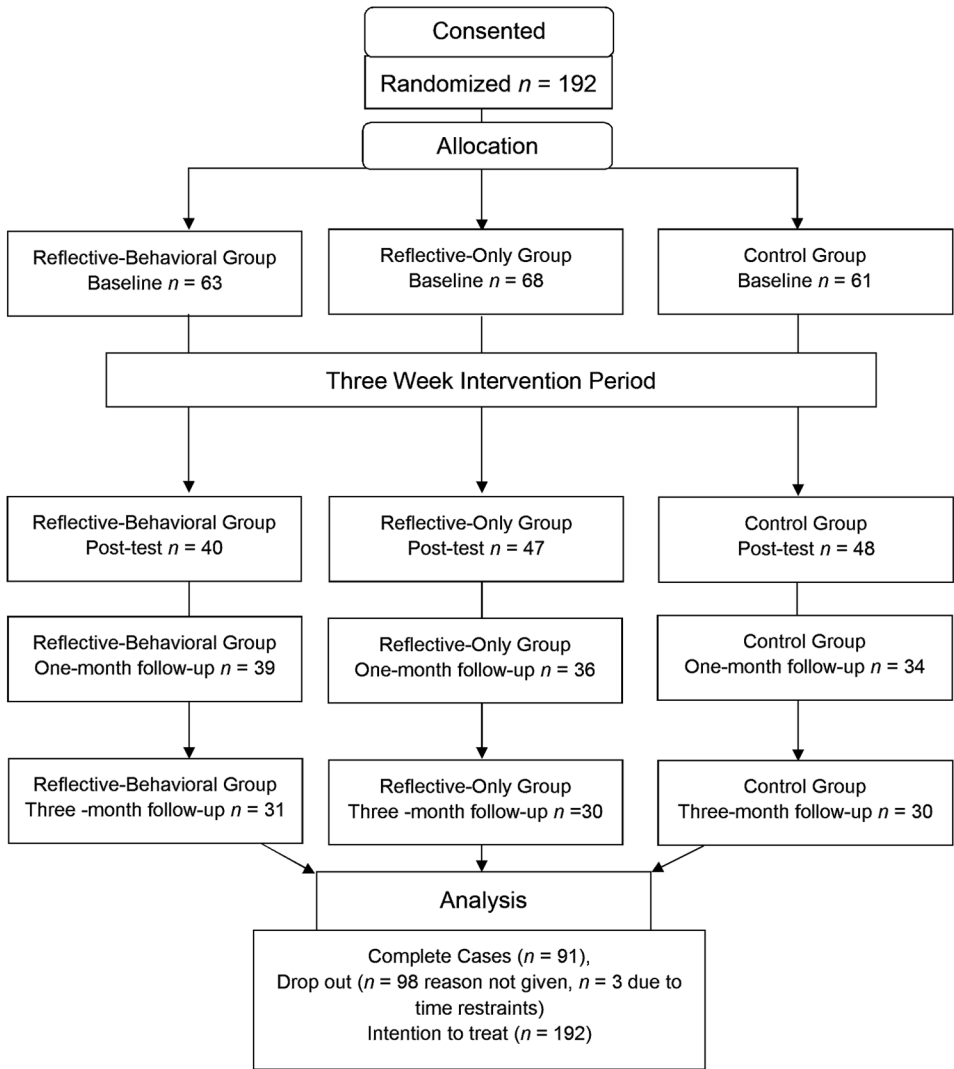
Figure 1.   Trial accrual and retention adapted from the CONSORT statement (Schulz et al., 2010).

There were no differences between the reflective-only and reflective-behavioral intervention regarding who they directed, or focused their gratitude on, $\chi 2(4, n = 61) = .422, p = 0.981$, with the majority of participants (62.3%) writing about a "mixture of all" different people and relationships, (i.e., family members, nonfamily members, friends, nonfamily members, boyfriend/girlfriend). There were no significant association between intervention assignment and enjoyment of the intervention task, $F(2,126) = 0.43, p = 0.655$, or how much they like writing in general, $F(2,126) = 1.50, p = 0.226$.

Changes in Psychosocial Variables Between Interventions Over Time

Changes in psychosocial variables over time between the interventions are depicted in Table 2.

Effects of condition at immediate posttest. Mixed repeated measures analyses of variance revealed a statistically significant Intervention × Time (baseline, immediate posttest) interaction for scores on affect balance, $F(2, 118) = 3.51, p = .03, \eta^2_p = .06$, and depression,

Table 2
Changes in Psychosocial Variables Across Time Between Intervention Conditions

| | Within-subject effects (Baseline, posttest) | | | | Within-subject effects (Baseline, posttest, 1 month) | | | | Within-subject effects (Baseline, posttest, 1 month, 3 months) | | | |
| | Time | | Interaction x Time | | Time | | Interaction x Time | | Time | | Interaction x Time | |
	F	p	F	p	F	p	F	p	F	p	F	p
Life satisfaction	2.25	.14	1.11	.33	3.56	.03*	.84	.50	2.30	.08	.80	.57
Affect balance	3.23	.08	3.51	.03*	3.56	.03*	2.59	.04*	2.07	.11	2.70	.02*
Positive affect	.70	.40	2.14	.12	.12	.89	.80	.53	.41	.75	1.67	.13
Negative affect	10.22	.002**	2.31	.10	8.88	<.001***	2.77	.03*	4.49	.005**	2.13	.052
Depression	14.14	<.001***	3.57	.03*	7.54	<.001***	1.88	.12	4.24	.006**	1.63	.14
Gratitude	.004	.95	.91	.41	.29	.75	1.26	.29	1.40	.24	1.43	.21
Gratitude Expression	1.37	.24	.05	.95	1.04	.36	.14	.97	.90	.44	2.50	.02*
Relationship Satisfaction	.01	.91	1.27	.29	.41	.67	1.12	.35	.54	.65	1.56	.16

Note. F for time and interaction effects (general linear model repeated measures).

*$p < .05$. **$p < .01$. ***$p < .001$.

$F(2, 115) = 3.57$, $p = .03$, $\eta^2_p = .06$. Results showed no significant Intervention x Time (baseline, immediate posttest) interactions for scores on life satisfaction, positive affect and negative affect subscales, gratitude, gratitude expression, or relationship satisfaction. There was, however, a significant main effect for time on negative affect scores, $F(1, 119) = 10.22$, $p = .002$, $\eta^2_p = .08$.

Post hoc analysis revealed statistically significant differences between the mean scores at baseline and follow-up (mean difference = 1.12, standard error [SE] = .35, $p = .002$, 95% CI [.426, 1.81]), indicating a steady decrease in negative affect across time for all conditions, with those in the reflective-behavioral intervention showing the greatest decreases from baseline to immediate posttest (mean difference = 1.97). Post hoc pairwise comparisons on significant interactions revealed that participants in the reflective-behavioral intervention had significantly reduced depression scores from baseline to immediate posttest (mean difference = 2.97, $SE = .73$, $p < .001$, 95% CI [1.52, 4.42]) and significant increases in affect balance from baseline to immediate posttest (mean difference = −2.30, $SE = .97$, $p = .019$, 95% CI [−4.22, −.39]). No such difference were found in depression or affect balance for either the reflective-only intervention ($p > .05$) or the control intervention ($p > .05$).

Sensitivity analysis using multilevel growth modeling was conducted on the significant findings, using all available data ($n = 192$) with an ITT approach. For affect balance as outcome, a basic model in which all parameters were fixed produced a -2LL of 1994.40. Including a random intercept to the model significantly improved the overall fit of the model, -2LL = 1938.99, $\chi^2(1) = 55.41$, $p < 0.01$. However, inclusion of a random slope, allowing covariance between the random slopes and random intercepts or allowing time to be repeated did not significantly improve model fit. Therefore, the second model with random intercepts and eight parameters was used for the multilevel analysis on affect balance. Omnibus tests of significance for the predictor variables included in the model supported mixed repeated measures results and revealed a significant treatment x time interaction, $F(2, 144.54) = 3.22$, $p = .046$, indicating that trends in affect balance varied significantly over time across the three assigned interventions.

The maximum likelihood estimates of the fixed effects parameters revealed evidence of differing trends over time between the control condition (reference category) and the reflective-behavioral condition that was statistically significant (estimate = 2.90, $SE = 1.20$, $t = 2.41$, $p = .017$, 95% CI [.522, 5.27]). This estimated that the slope from baseline to immediate posttest for participants in the reflective-behavioral intervention was 2.90 points greater than the slope for those in the control condition. There was also evidence of differing trends over time between the reflective-only intervention and the control intervention that approached significance, but it was not statistically significant (estimate = 1.96, $SE = 1.14$, $t = 1.71$, $p = .089$, 95% CI [−.304, 4.22]) and estimated that the slope from baseline to immediate posttest for participants in the reflective-only intervention was .21 points greater than the slope for those in the control intervention.

Further, there was no evidence of differing trends over time between the reflective-behavioral intervention and the reflective-only intervention (reference category; estimate = .933, $SE = 1.20$, t = 1.77, p = .439, 95% CI [−1.44, 3.31]). Separate models developed for each of the conditions estimated that participants in the reflective-behavioral intervention indicated the largest changes in affect balance over time. This average rate of change over time was approaching statistical significance, beginning with an average affect balance score of 6.84 at baseline and increasing by 1.95 points on average at immediate posttest (estimate = 1.95, $SE = 1.00$, $t = 1.95$, $p = 0.057$, 95% CI [−.061, 3.97]). The average rate of change over time was not statistically significant for participants in the reflective-only intervention (estimate = 1.01, $SE = .83$, $t = 1.22$, $p = .229$, 95% CI [−.656, 2.67]) or the control intervention (estimate = −.903, $SE = .67$, $t = −1.35$, $p = .184$, 95% CI [−2.25, .442]).

For depression as outcome, a basic model, in which all parameters were fixed, produced a -2LL of 1880.36. Including a random intercept to the model significantly improved the overall fit of the model, -2LL = 1821.30, $\chi^2(1) = 59.06$, $p < 0.01$. However, inclusion of a random slope allowing covariance between the random slopes and random intercepts or allowing time to be repeated did not significantly improve model fit. Therefore, the second model with random intercepts and eight parameters was used for the multilevel analysis on depression. Omnibus

tests of significance for the predictor variables included in the model partially supported mixed repeated measures results and revealed a marginally but not statistically significant treatment x time interaction, $F(2, 148.15) = 2.64$, $p = .075$, indicating that trends in depression varied over time across the three assigned interventions.

The maximum likelihood estimates of the fixed effects parameters revealed evidence of significantly differing trends over time between the control (reference category) and the reflective-behavioral intervention that was statistically significant (estimate $= -1.85$, $SE = .87$, $t = -2.126$, $p = .035$, 95% CI [-3.56, $-.131$]). This estimated that the slope from baseline to immediate posttest for participants in the reflective-behavioral intervention was 1.85 points less than the slope for those in the control condition. There was no evidence of differing trends over time between the reflective-only and the control intervention (estimate $= -.19$, $SE = .82$, $t = -.233$, $p = .816$, 95% CI [-1.82, 1.43]); however, there was evidence of differing trends over time between the reflective-behavioral and the reflective-only intervention (reference category) that approached significance, but it was not statistically significant (estimate $= -1.66$, $SE = .87$, $t = -1.90$, $p = .059$, 95% CI [-3.38, $.067$]).

Separate models developed for each of the conditions estimated that participants in the reflective-behavioral intervention indicated the largest changes in depression over time. This average rate of change over time was statistically significant, beginning with an average depression score of 12.41 at baseline and decreasing by 2.51 points on average at immediate posttest (estimate $= -2.51$, $SE = .74$, $t = -3.37$, $p = 0.001$, 95% CI [-4.00, -1.01]). The average rate of change over time was not statistically significant for participants in the reflective-only (estimate $= -.863$ $SE = .52$ $t = -1.66$, $p = .103$, 95% CI [-1.91, $.180$]) or the control intervention (estimate $= -.655$, $SE = .56$, $t = -1.18$, $p = .25$, 95% CI [-1.77, $.463$]). The superiority of the reflective-behavioral intervention in improving affect balance and reducing depressive symptoms from baseline to immediate posttest echoed results from the mixed repeated measures.

Effects of condition at 1-month follow-up. A mixed repeated measures analysis of variance revealed a statistically significant Intervention $\times$ Time (baseline, immediate posttest, 1-month) interaction for scores on affect balance, $F(4, 158) = 2.59$, $p = .039$, $\eta^2_p = .06$, and the negative affect subscale, $F(4, 158) = 2.77$, $p = .029$, $\eta^2_p = .07$. Results showed no significant Intervention x Time (baseline, immediate posttest, 1-month) interactions for scores on life satisfaction, positive affect subscale, depression, gratitude, gratitude expression, or relationship satisfaction. There was however, a significant main effect for time on life satisfaction scores, $F(2, 164) = 3.56$, $p = .03$, $\eta^2_p = .04$.

Post hoc analysis revealed statistically significant differences between the mean scores at baseline and follow-up (mean difference $= -1.11$, $SE = .435$, $p = .013$, CI [-1.98, $-.242$]), indicating a steady increase in life satisfaction scores across time for all conditions, with those in the reflective-behavioral intervention showing the greatest increases from baseline to immediate posttest (mean difference $= -1.89$) and follow-up (mean difference $= -1.64$). Post hoc pairwise comparisons on significant interactions revealed that participants in the reflective-behavioral intervention had significant increases in affect balance from baseline to immediate posttest (mean difference $= 2.88$, $SE = 1.17$, $p = .016$, 95% CI [-5.22, $-.544$]) and marginally significant increases from baseline to 1-month follow-up (mean difference $= -2.16$, $SE = 1.18$, $p = .071$, 95% CI [14.512, .192]); they had significant decreases in negative affect scores from baseline to immediate posttest (mean difference $= 2.60$, $SE = .79$, $p = .002$, 95% CI [1.03, 4.17]) and from baseline to 1-month follow-up (mean difference $= 1.88$, SE $= .75$, $p = .014$, 95% CI [.384, 3.38]).

Participants in the reflective-only intervention had significant increases in affect balance from baseline to follow-up (mean difference $= -2.85$, $SE = 1.16$, $p = .016$, 95% CI [-5.15, $-.539$]); they had significant decreases in negative affect scores from baseline to immediate posttest (mean difference $= 1.77$, $SE = .78$, p $= .025$, 95% CI [.226, 3.312]) and baseline to 1-month follow-up (mean difference $= 2.39$, $SE = .74$, p $= .002$, 95% CI [.918, 3.85]). No such differences were found in affect balance or negative affect for the control intervention (*ps* $> .05$). These significant findings did not hold for the sensitivity analysis using an ITT approach (See Table 3).

Table 3
Linear Mixed Modeling Fixed Effects of Treatment by Time

Time points	Model parameters	Model fit (-2LL)	Type III tests of fixed effects: treatment x time	
T1,T2			*F*	*p*
Affect balance	8, random intercepts	1938.99	3.22	.043
Depression	8, random intercepts	1821.30	2.64	.075
T1, T2, T3:				
Affect balance	9, random intercepts & slopes	2556.77	2.35	.098
Negative affect	8, random intercepts	2223.54	1.55	.214
T1, T2, T3, T4:				
Affect balance	9, random intercepts & slopes	3119.06	.35	.70
Negative affect	8, random intercepts	2700.65	.63	.531
Gratitude Expression	7, random intercepts	2066.73	1.92	.148

Effects of condition at 3-month follow-up. A mixed repeated measures analysis of variance revealed a statistically significant Intervention × Time (baseline, immediate posttest, and 1-month and 3-month follow-up) interaction for scores on affect balance, $F(6, 168) = 2.70$, $p = .016$, $\eta^2_p = .09$, scores on the negative affect subscale, $F(6, 168) = 2.13$, $p = .052$, $\eta^2_p = .07$, and gratitude expression, $F(6, 168) = 2.50$, $p = .024$, $\eta^2_p = .08$. Results showed no significant Intervention x Time (baseline, immediate posttest, and 1-month and 3-month follow-up) interactions for scores on life satisfaction, positive affect subscale, depression, gratitude, gratitude expression, or relationship satisfaction. There was however, a significant main effect for time on depression scores, $F(3, 171) = 4.24$, $p = .006$, $\eta^2_p = .07$.

Post hoc analysis revealed statistically significant differences between the mean scores at baseline and immediate posttest (mean difference = 1.73, $SE = .52$, $p = .001$, 95% CI [.699, 2.77]) and baseline to 1-month follow-up (mean difference = 1.54, $SE = .50$, $p = .003$, 95% CI [.533, 2.55]), but not from baseline to 3-month follow-up (mean difference = .872, $p > .05$), indicating a steady decrease in depression scores for up to 1 month after intervention cessation. Those in the reflective-behavioral intervention showed the greatest decrease from baseline to immediate posttest (mean difference = 3.74), 1-month follow-up (mean difference = 1.9), and 3-month follow-up (mean difference = 1.63).

Post hoc pairwise comparisons on significant interactions revealed that participants in the reflective-behavioral intervention had significant increases in affect balance from baseline to immediate posttest (mean difference = −.358, $SE = 1.32$, $p = .009$, 95% CI [−6.22, −.937]) and marginally but not statistically significant increases from baseline to 1-month follow-up (mean difference = −2.53, $SE = 1.43$, $p = .082$, 95% CI [−5.39, .333]); there were no significant changes from baseline to 3-month follow-up ($p = .916$). Similarly, participants in the reflective-only intervention had marginally significant increases in affect balance from baseline to immediate posttest (mean difference = −2.45, $SE = 1.29$, $p = .062$, 95% CI [−5.03, .125]) and statistically significant increases from baseline to 1-month follow-up (mean difference = −3.05, $SE = 1.39$, $p = .033$, 95% CI [−5.84, −.263]); there were no significant changes from baseline to 3-month follow-up ($p = .359$). No such improvements in affect balance were found for participants in the control intervention, who actually experienced initial reductions in affect balance from baseline to immediate posttest (mean difference = 2.65, $SE = 1.29$, $p = .044$, 95% CI [.075, 5.23]) and then returned back to baseline levels at 1-month ($p = .520$) and 3-month follow-up ($p = .475$).

Participants in the reflective-behavioral intervention had significant decreases in negative affect from baseline to immediate posttest (mean difference = 3.05, $SE = .933$, $p = .002$, 95% CI [1.18, 4.92]) and 1-month follow-up (mean difference = 2.26, $SE = .904$, $p = .015$, 95% CI [.453, 4.07]); there were no significant changes from baseline to 3-month follow-up ($p = .418$). Similarly, participants in the reflective-only intervention had significant decreases in negative affect from baseline to immediate posttest (mean difference = 2.15, $SE = .91$, $p = .022$, 95% CI

Table 4

Simple Mediation Model: Affect Balance T1–T3 Mediated by Gratitude Expression

Model	Estimate	SE	p	BCa 95% CI (lower)	BCa 95% CI (upper)
Model without Mediator					
Intercept	7.65	5.30	0.159	−3.15	18.46
AffectBT1 T3(c)	0.761	0.156	<0.0001	0.444	1.078
[controlling for GratExT1]					
R^2(y,x)	0.429	–	<0.001		
Model with Mediator					
Model 1:GratExT3 outcome					
Intercept	8.94	1.32	<0.0001	6.25	11.63
AffectBT1 GratExT3(a)	0.08	0.039	<0.05	0.001	0.159
Model 2:AffectBT3 outcome					
Intercept	−.939	7.39	0.214	−24.47	5.69
GratExT3 AffectBT3(b)	1.906	0.635	<0.01	0.612	3.201
AffectBT1 AffectBT3(c')	0.609	0.148	<.001	0.307	0.911
Indirect Effect (ab)	0.152	0.098	–	.015	.435
R^2(m,x)	0.276	–	<0.01		
R^2(y,m,x)	0.558	–	<0.0001		

Note. SE = standard error; BCa 95% CI= bias-corrected and accelerated 95% confidence interval with a resample procedure of 5,000 bootstrap samples. AffectBT1 = affect balance at time 1, AffectBT3 = affect balance at time 3, GratExT1 = gratitude expression at time 1, GratExT3 = gratitude expression at time 3.

[.328, 3.97]) and 1-month follow-up (mean difference = 2.35, *SE* = .88, *p* = .01, 95% CI [5.85, 4.12]); there were no significant changes from baseline to 3-month follow-up (*p* = .247). No such decreases in negative affect were found for participants in the control intervention (all *ps* > .2).

Unexpectedly, participants in the reflective-behavioral and reflective-only intervention had no significant increases in gratitude expression from baseline to immediate posttest, 1-month follow-up, or 3-month follow-up (all *ps* > .1); however, there were significant increases in gratitude expression for the control intervention from baseline to 3-month follow-up only (mean difference = −.89, *SE* = .42, p = .038, 95% CI [−1.73, −.05]). Analysis of variance was conducted to explore this further and indicated that there were no differences between the intervention groups on gratitude expression scores at 3-month follow-up, $F(2, 87) = .467$, *p* = .629. These significant findings did not hold for the sensitivity analysis using an ITT approach (see Table 3).

Testing for Mediated Relationships on Findings

Participants who completed the reflective-behavioral gratitude intervention showed changes in affect balance from baseline (T1) to 1-month follow-up (T3). As such, a simple mediation analysis was conducted to examine whether this effect was mediated by levels of gratitude expression (T3; controlling for T1). None of the model variables exceeded the recommended intercorrelation value of > 0.80 (Katz, 2011); tolerance values were all above 0.6 and variance inflation factors values were < 10, signifying a low risk of multicollinearity in the data (Field, 2013).

This analysis confirmed that there was a significant total effect of affect balance from T1 to T3, controlling for T1 gratitude expression scores (*c*), and this positive association was reduced when the effect of gratitude expression (T3) was also taken into account (*c'*). Gratitude expression (T3) mediated the relationship between T1 and T3 affect balance, as indicated by a significant indirect effect (*ab*). The data suggest that for those who completed the reflective-behavioral gratitude intervention, affect balance at T1 leads to greater levels of gratitude expression in relationships, which in turn positively affects affect balance at T3. Overall, this model significantly accounted for 55.8% of the variance in affect balance at T3, $F(3, 31) = 13.05$ *p* < .0001. See Table 4 for all parameter estimates.
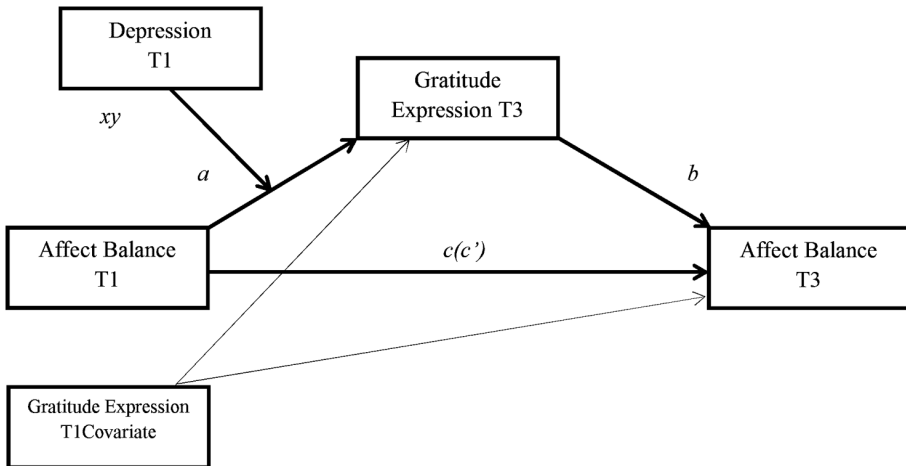
Figure 2.   Moderated mediation model for participants in the reflective-behavioral gratitude condition.

To test the robustness of our results, an alternative model examining whether the effect of T1 and T3 affect balance was mediated by levels of overall gratitude (T3; controlling for T1), rather than expressed gratitude, was conducted and results did not support this model. Although the indirect effect of affect balance from T1 to T3 mediated by overall gratitude was indeed positive ($B = .115$), this was not statistically significant (95% BCa CI [$-.055, .390$]). This indicates that it is the expression of gratitude rather than overall levels of gratitude that drives this effect.

In further support of this interpretation, for participants in the reflective-only gratitude intervention, the above mediation pathway via gratitude expression was not statistically significant ($B = .019$, 95% BCa CI [$-.032, .25$]), as evidenced by a nonsignificant indirect effect. Furthermore, for these participants, overall gratitude levels did not account for or mediate the relationship between T1 and T3 affect balance, ($B = -.02$, 95% BCa CI [$-.264, .093$]), as evidenced by a nonsignificant indirect effect. Finally, none of the above-mentioned mediation models were significant when examining changes in affect balance from baseline to T2 or from baseline to T4.

Testing for Moderated Mediation on Significant Findings

For participants who completed the reflective-behavioral gratitude intervention, changes in affect balance from T1 to T3 was mediated by increases in gratitude expression. Consequently whether this mediated effect differed depending on participant's baseline depressive scores was examined using moderated mediation analysis, represented in Figure 2. Results indicated that the pathway between T1 affect balance and T3 gratitude expression was moderated by depression at T1 (xy; $B = .019$, 95% BCa CI [$.005, .033$]). This interaction and the simple slopes were probed at the mean (9.5), 1 SD above the mean (14.16), and 1 SD below the mean (4.84) of the depression scores. The results showed evidence of a significant indirect effect of T1 affect balance on T3 affect balance through gratitude expression (T3; controlling for T1) for participants with higher levels of depression at T1 ($B = .22$, 95% BCa CI [$.022, .558$]), but not for participants with average ($B = .06$, 95% BCa CI [$-.073, .263$]) or lower levels of depression ($B = -.103$, 95% BCa CI [$-.39, .16$]). Therefore, the change from T1 and T3 affect balance was mediated by improvements in gratitude expression, but this effect was conditional upon participants T1 levels of depression.

Discussion

The Efficacy of the Interventions

The primary aim of this study was to compare the effects of two interpersonal gratitude journals, one with a reflective component only and the other with both a reflective and a

behavioral component compared to a neutral control journal, on several psychosocial outcomes over 3 months. Analysis of efficacy using complete cases showed that those who completed the reflective-behavioral gratitude intervention experienced a significant increase in affect balance and a decrease in depressive symptoms at immediate posttest, with no such differences found for either the reflective-only gratitude intervention or the control intervention. Sensitivity analysis using all randomized participants replicated these findings. Therefore, interpersonal other-oriented gratitude appears to be particularly effective in enhancing emotional well-being when this gratitude is outwardly expressed.

To our knowledge, this is the first longitudinal randomized controlled trial to show that gratitude, tailored specifically to others, which is felt (reflective) and expressed (behavioral), rather than grateful reflection alone, causes greater beneficial change. At 1-month follow-up, participants in both gratitude interventions experienced decreases in negative affect and improvements in affect balance, with no changes for participants in the control intervention. However, these findings at 1-month follow-up were not supported by the sensitivity analysis. At 3-month follow-up, all outcomes began returning to baseline levels. Unexpectedly, participants in the control intervention experienced increases in gratitude expression at 3-month follow-up. We speculate that perhaps the surveys at each follow-up served as a self-regulation prompt and these participants began to self-regulate their gratitude to some extent (Sitzmann & Katherine, 2010). Having said that, these findings were not maintained using sensitivity analysis.

In sum, those who completed the reflective-behavioral gratitude journal experienced the most improvements in outcomes at immediate posttest. However, in the month that followed the intervention, the reflective-only gratitude intervention performed as well as the reflective-behavioral gratitude intervention; at 3-month follow-up, no differences were observed. This supports the findings that continued practice (Seligman et al., 2005) or engagement in varied PPIs may be needed to curtail this hedonic adaptation (Layous & Lyubomirsky, 2014). Nonetheless, these small to medium effect sizes, given the relatively brief nature of the intervention and rigorous methodology employed, warrant further investigation before any conclusive recommendations regarding optimum use of gratitude interventions as a valuable therapeutic agent can be offered.

Mediating and Moderating Factors

Examining underlying mechanisms in the reflective-behavioral gratitude intervention showed evidence that changes in affect balance from baseline to 1-month follow-up partially accounted for the extent to which people expressed gratitude within their social relationships at 1-month follow-up, controlling for preintervention levels. This provides evidence that it was the expression of gratitude that in part accounted for the positive changes in affect balance from baseline to 1-month follow-up, as predicted. Overall levels of gratitude did not mediate this relationship, and the above-mentioned pathway was only evident for participants in the reflective-behavioral gratitude intervention. This is the first study to demonstrate that it is through the process of expressing gratitude to others that a reflective-behavioral gratitude journal affects well-being.

Additionally, this mediated effect was most pronounced for participants with higher level of depression at baseline. This is consistent with and gives strength to previous findings (Seligman et al., 2005; Sin & Lyubomirsky, 2009), suggesting that PPIs are not limited to the healthy. Reflecting the rationale of Froh and colleagues (2009), it may be the case that participants higher in depressive symptoms are likely to experience gratitude much less frequently, and as such this grateful reflection and expression is a more novel experience for them and may be what they need to initiate positive social interactions. This lends support to the view that people experiencing mild depressive symptoms or "dysphoric" individuals are likely to garner the benefits from PPIs and gratitude interventions may be particularly effective for treating affective symptoms.

People's initial affective status appears to be of particular importance for future studies because this sample had near ceiling levels of gratitude before engaging in the intervention; this is also characteristic of past studies (see Wood et al., 2010). Thus, it is recommended that future researchers target subclinical and clinical populations, as well as people who present with less than desirable levels of the prespecified distal and proximal outcomes of the intervention.

By preassessing for ceiling effects, researchers can focus on people who appear to be in need of a gratitude-based intervention, and there is therefore greater potential for both change to occur and for improving that person's mental health. As gratitude interventions are becoming increasingly widespread as a therapeutic/clinical technique in a range of applied milieus (Davis et al., 2016; Emmons & Stern, 2013; Kerr et al., 2015; Rashid, 2015; Seligman, Rashid, & Park, 2006; Wong et al., 2016), this may be of crucial importance for practitioners in clinical psychology settings who seek to employ gratitude interventions to improve client mental health and well-being.

Limitations

Although there were no significant differences between participants who remained for the duration of the study and those who dropped out on variables that were assessed, 1-month follow-up changes did not hold under sensitivity analysis using an ITT approach. Therefore, we remain cautious about the effectiveness of these interventions in the weeks after the intervention ends. This being said, this sensitivity analysis is considered very conservative and more susceptible to Type II error in that it ignores noncompliance, protocol deviations, and drop out (Fergusson, Aaron, Guyatt, & Hébert, 2002; Gupta, 2011; Hollis & Campbell, 1999). This is particularly conservative and problematic when there is a large amount of missing responses, like in the present study. As such, it appears that the benefits of the intervention were found only for those who complied with the treatment received and adhered to the protocol and thus represent the optimal case treatment results that could be achieved. It is argued that ITT approaches may serve as a test of treatment allocation rather than the actual treatment per se (Petkova & Teresi, 2002), and a more suitable application of the ITT approach is achievable when complete outcome data are available for all participants who were randomized.

Additionally, given the large number of dependent variables, there is an increased risk for Type 1 error. However, this potential concern is addressed by the prespecification of a single primary efficacy outcome in the trial protocol, the choice to run two-tailed tests, which are more conservative, and by conducting a sensitivity analysis using multilevel modeling and an ITT approach on any significant findings to determine their robustness. Although multiple techniques were adopted in the present study to reduce attrition and improve adherence (financial incentive, reminder text messages, providing participants with journals, option of completing follow-up by hand or online, etc.), at the final 3-month follow-up, less than half of the participants returned. High attrition rates are a common feature in PPI studies (Bolier et al., 2015; Mongrain & Anselmo-Matthews, 2005; Schueller & Parks, 2012).

Drawing on the positive activity model (Lyubomirsky & Layous, 2013) and past findings (Giannopoulosa & Vella-Brodrick, 2011), it is possible that many participants who dropped out may not have perceived a "fit" between themselves and the intervention, and as such there may have been little intrinsic motivation or interest for them to continue and complete follow-ups. In addition to more common retention methods, future research should continue to strive for lower drop-out through taking into account the intersection between the PPI and the individual that might produce an optimal "person activity fit."

In light of previous research (O'Connell et al., 2016a; O'Connell et al., 2016), it is surprising that these gratitude interventions, designed to foster social relationships, did not significantly improve participant's levels of gratitude, life satisfaction, or relationship satisfaction, relative to the neutral control intervention. It is possible that because of the large concentration of scores near the upper limits of these measures, there may have been ceiling effects occurring, as seen in previous work of this nature (Davis et al., 2016; Froh et al., 2009; Wood et al., 2010), so there is less potential for improvements to occur. Further, many intervention studies do not find significant improvements in gratitude compared to measurement-only control groups (Davis et al., 2016; Senf & Liau, 2013). Future researchers are advised to examine the efficacy and effectiveness of these interventions on less "emotionally healthy" samples.

Conclusion

The present outcomes support the growing evidence that gratitude interventions can improve markers of well-being while highlighting that these effects are not always straightforward and

sustainable. As suggested in a recent meta-analysis (Davis et al., 2015), booster sessions may be needed to achieve longer term sustainable improvements in well-being and avoid so-called hedonic adaption. These findings offer novel evidence that expressing felt gratitude to others may be a key step in reaping postintervention benefits. These interventions were very straightforward and cost-effective, and with the rise in the application of gratitude interventions in practice, this research offers unique understanding in terms of not only *if* but also *how* gratitude interventions exert effects. Importantly, this was accomplished through employing robust high-quality randomized controlled design and pretrial registration. These advancements are needed in future positive psychology intervention research to address past critiques and increase legitimacy in the health, occupational, and clinical research community.

References

Algoe, S. B., Haidt, J., & Gable, S. L. (2008). Beyond reciprocity: Gratitude and relationships in everyday life. Emotion, 8(3), 425–429.

Andresen, E. M., Carter, W. B., Malmgren, J. A., & Patrick, D. L. (1994). Screening for depression in well older adults: Evaluation of a short form of the CES-D. American Journal of Preventative Medicine, 10, 77–84.

Boehm, J. K., Lyubomirsky, S., & Sheldon, K. M. (2011). A longitudinal experimental study comparing the effectiveness of happiness-enhancing strategies in Anglo Americans and Asian Americans. Cognition & Emotion, 25(7), 1263–1272.

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(1), 119

Briner, R. B., & Walshe, N. D. (2015). An evidence-based approach to improving the quality of resource-oriented well-being interventions at work. Journal of Occupational and Organizational Psychology, 88(3), 563–586.

Clauss, E., Hoppe, A., O'Shea, D., Michel, A., Gonzalez-Morales, M. G., & Steidle, A. (in press). Promoting personal resources and reducing exhaustion through positive reflection among caregivers. Journal of Occupational Health Psychology

Davis, D. E., Choe, E., Meyers, J., Wade, N., Varjas, K., Gifford, A., & Worthington, E. L. Jr. (2016). Thankful for the little things: A meta-analysis of gratitude interventions. Journal of Counseling Psychology, 63, 20–31.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. Journal of Personality Assessment, 49(1), 71–75.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. Social Indicators Research, 97(2), 143–156.

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(2), 377–389.

Emmons, R. A., & Mishra, A. (2011). Why gratitude enhances well-being: What we know, what we need to know. In K. M. Sheldon, T. B. Kashdan, & M. F. Steger (Eds.), Designing positive psychology: Taking stock and moving forward (pp 248–262). New York: Oxford University Press.

Emmons, R. A., & Stern, R. (2013). Gratitude as a psychotherapeutic intervention. Journal of Clinical Psychology, 69(8), 846–855.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41, 1149–1160.

Fergusson, D., Aaron, S. D., Guyatt, G., & Hébert, P. (2002). Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. BMJ, 325(7365), 652–654.

Field, A. (2013). Discovering statistics using IBM SPSS statistics. Thousand Oaks, CA: Sage.

Fredrickson, B.L., 2004. The broaden-and-build theory of positive emotions. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences (359), 1367–1378.

Froh, J. J., Kashdan, T. B., Ozimkowski, K. M., & Miller, N. (2009). Who benefits the most from a gratitude intervention in children and adolescents? Examining positive affect as a moderator. The Journal of Positive Psychology, 4(5), 408–422.

Giannopoulos, V. L., & Vella-Brodrick, D. A. (2011). Effects of positive interventions and orientations to happiness on subjective well-being. The Journal of Positive Psychology, 6(2), 95–105.

Gordon, C. L., Arnette, R. A., & Smith, R. E. (2011). Have you thanked your spouse today?: Felt and expressed gratitude among married couples. Personality and Individual Differences, 50(3), 339–343.

Gupta, S. K. (2011). Intention-to-treat concept: A review. Perspectives in Clinical Research, 2(3), 109–112.

Hayes, A. F. (2013). An introduction to mediation, moderation, and conditional process analysis. New York: The Guilford Press.

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ, 319(7211), 670–674.

Irwin, M., Artin, K. H., & Oxman, M. N. (1999). Screening for depression in the older adult: Criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). Archives of Internal Medicine, 159(15), 1701–1704.

Katz, M. H. (2011). Multivariable analysis: A practical guide for clinicians and public health researchers. New York: Cambridge University Press.

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(1), 17–36.

Kong, F., Ding, K., & Zhao, J. (2015). The relationships among gratitude, self-esteem, social support and life satisfaction among undergraduate students. Journal of Happiness Studies, 16(2), 477–489.

Krejtz, I., Nezlek, J. B., Michnicka, A., Holas, P., & Rusanowska, M. (2016). Counting one's blessings can reduce the impact of daily stress. Journal of Happiness Studies, 17(1), 25–39.

Lambert, N. M., Clark, M. S., Durtschi, J., Fincham, F. D., & Graham, S. M. (2010). Benefits of expressing gratitude expressing gratitude to a partner changes one's view of the relationship. Psychological Science 21(4), 574–580.

Lambert, N. M., & Fincham, F. D. (2011). Expressing gratitude to a partner leads to more relationship maintenance behavior. Emotion, 11(1), 52–60.

Layous, K., Chancellor, J., & Lyubomirsky, S. (2014). Positive activities as protective factors against mental health conditions. Journal of Abnormal Psychology, 123(1), 3–12.

Layous, K., Chancellor, J., Lyubomirsky, S., Wang, L., & Doraiswamy, P. M. (2011). Delivering happiness: Translating positive psychology intervention research for treating major and minor depressive disorders. The Journal of Alternative and Complementary Medicine, 17(8), 675–683.

Layous, K., & Lyubomirsky, S. (2014). The how, why, what, when, and who of happiness: Mechanisms underlying the success of positive activity interventions. Positive Emotion: Integrating the Light Sides and Dark Sides, 473–495.

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? Psychological Bulletin,131(6), 803–855.

Lyubomirsky, S., & Layous, K. (2013). How do simple positive activities increase well being? Current Directions in Psychological Science, 22(1), 57–62.

McCullough, M. E., Emmons, R. A., & Tsang, J. A. (2002). The grateful disposition: A conceptual and empirical topography. Journal of Personality and Social Psychology, 82(1), 112–127.

Mohr, D. C., Spring, B., Freedland, K. E., Beckner, V., Arean, P., Hollon, S. D., & Kaplan, R. (2009). The selection and design of control conditions for randomized controlled trials of psychological interventions. Psychotherapy and Psychosomatics, 78, 275–284.

Mongrain, M., & Anselmo-Matthews, T. (2012). Do positive psychology exercises work? A replication of Seligman et al.(2005). Journal of Clinical Psychology, 68(4), 382–389.

O'Connell, B. H., O'Shea, D., & Gallagher, S. (2016a). Enhancing social relationships through positive psychology activities: a randomised controlled trial. The Journal of Positive Psychology, 1–14.

O'Connell, B. H., O'Shea, D., & Gallagher, S. (2016b). Mediating effects of loneliness on the gratitude-health link. Personality and Individual Differences, 98, 179–183.

O'Connell, B. H., O'Shea, D., & Gallagher, S. (2016c). Examining psychosocial pathways underlying gratitude interventions: a randomized controlled trial. Manuscript submitted for publication.

Parks, A. C. (2014). A case for the advancement of the design and study of online positive psychological interventions. The Journal of Positive Psychology, 9(6), 502–508.

Parks, A. C., Della Porta, M.D., Pierce, R. S., Zilca, R., & Lyubomirsky, S. (2012). Pursuing happiness in everyday life: The characteristics and behaviors of online happiness seekers. Emotion, 12(6), 1222–1234.

Parks, A. C., & Schueller, S. M. (Eds.). (2014). The Wiley-Blackwell handbook of positive psychological interventions. New York: Wiley Blackwell.

Pavot, W., & Diener, E. (1993). Review of the satisfaction with life scale. Psychological Assessment, 5(2), 164–172.

Petkova, E., & Teresi, J. (2002). Some statistical issues in the analyses of data from longitudinal studies of elderly chronic care populations. Psychosomatic Medicine, 64(3), 531–547.

Proyer, R. T., Wellenzohn, S., Gander, F., & Ruch, W. (2015). Toward a better understanding of what makes positive psychology interventions work: Predicting happiness and depression from the person × intervention fit in a follow-up after 3.5 years. Applied Psychology: Health and Well-Being, 7(1), 108–128.

Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. Applied Psychological Measurement, 1(3), 385–401.

Rash, J. A., Matsuba, M. K., & Prkachin, K. M. (2011). Gratitude and Well-Being: Who Benefits the Most from a Gratitude Intervention? Applied Psychology: Health and Well-Being, 3(3), 350–369.

Rashid, T. (2015). Positive psychotherapy: A strength-based approach. The Journal of Positive Psychology, 10(1), 25–40.

Redwine, L. S., Henry, B. L., Pung, M. A., Wilson, K., Chinh, K., Knight, B., & Mills, P. J. (2016). Pilot randomized study of a gratitude journaling intervention on heart rate variability and inflammatory biomarkers in patients with stage b heart failure. Psychosomatic Medicine, 78(6), 667–676.

Schramm, D. G., Marshall, J. P., Harris, V. W., & Lee, T. R. (2005). After "I do": The newlywed transition. Marriage & Family Review, 38(1), 45–67.

Schueller, S. M., & Parks, A. C. (2012). Disseminating self-help: Positive psychology exercises in an online trial. Journal of Medical Internet Research,14(3), e63.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. Annals of Internal Medicine, 152, 726–732.

Seligman, M. E., & Csikszentmihalyi, M. (2014). Positive psychology: An introduction (pp. 279–298). Springer Netherlands.

Seligman, M. E., Rashid, T., & Parks, A. C. (2006). Positive psychotherapy. American psychologist, 61(8), 774–788.

Seligman, M. E. P., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. American Psychologist, 60, 410–421. Senf, K., & Liau, A. K. (2013). The effects of positive interventions on happiness and depressive symptoms, with an examination of personality as a moderator. Journal of Happiness Studies, 14(2), 591–612.

Sheldon, K. M., & Lyubomirsky, S. (2006). How to increase and sustain positive emotion: The effects of expressing gratitude and visualizing best possible selves. The Journal of Positive Psychology, 1(2), 73–82.

Sin, N. L., Della Porta, M. D., & Lyubomirsky, S. O. N. J. A. (2011). Tailoring positive psychology interventions to treat depressed individuals. Applied Positive Psychology: Improving Everyday Life, Health, Schools, Work, and Society, 79–96.

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(5), 467–487.

Sitzmann, T., & Ely, K. (2010). Sometimes you need a reminder: The effects of prompting self-regulation on regulatory processes, learning, and attrition. Journal of Applied Psychology, 95(1), 132.

Wong, Y. J., Owen, J., Gabana, N. T., Brown, J. W., Mcinnis, S., Toth, P., & Gilman, L. (2016). Does gratitude writing improve the mental health of psychotherapy clients? Evidence from a randomized controlled trial. Psychotherapy Research, 1–11.

Wood, A. M., Froh, J. J., & Geraghty, A. W. (2010). Gratitude and well-being: A review and theoretical integration. Clinical Psychology Review, 30(7), 890–905.

Wood, A. M., Maltby, J., Gillett, R., Linley, P. A., & Joseph, S. (2008). The role of gratitude in the development of social support, stress, and depression: Two longitudinal studies. Journal of Research in Personality, 42(4), 854–871.