Web-Based Positive Psychology Interventions: A Reexamination of Effectiveness

Rosalind J. Woodworth, Angela O'Brien-Malone, Mark R. Diamond, and Benjamin Schüz

University of Tasmania

Objective: Seligman, Steen, Park, and Peterson (2005) suggested that positive psychology interventions (PPIs) contain specific, powerful, therapeutic ingredients that effect greater increases in happiness and reductions in depression than a placebo control. This study reexamined the three PPIs that Seligman et al. found to be most effective when delivered over the internet. **Method:** Three PPIs and a placebo control, identical with the interventions used by Seligman et al., were examined in a web-based, randomized assignment design. **Results:** Mixed-design analysis of variance and multilevel modeling showed that all interventions, including the placebo, led to significant increases in happiness and reductions in depression. The effects of PPIs were indistinguishable from those of the placebo control. **Conclusion:** Using web-based delivery, both PPIs and theoretically neutral placebos can increase happiness and reduce depression in self-selected populations. Possible explanations include that non-specific factors common to most therapeutic treatments are responsible for the observed changes, or that cultural or other context-related variables operate to account for the divergent findings. J. Clin. Psychol. 00:1–15, 2016.

Keywords: positive psychology; experimental replication; happiness; online therapy; signature strengths; e-mental health

In 2005, Seligman, Steen, Park, and Peterson (2005) reported the results of a large-scale randomized-assignment trial showing that, when delivered via the web, positive psychology interventions (PPIs) could increase participants' happiness and decrease their depression relative to the changes effected by a placebo control. On the basis that the interventions had been delivered using the Internet, rather than in a more traditional face-to-face therapeutic setting, Seligman et al. (2005) concluded that "powerful specific ingredients in the exercises" (p. 420) were responsible for the beneficial effects, rather than general factors related to the human therapeutic alliance.

However, in a review of five randomized controlled trials (including the trial by Seligman et al., 2005) that tested web-based positive psychology interventions, Mitchell, Vella-Brodrick, and Klein (2010) noted that only two (Mitchell, Stanimirovic, Klein, & Vella-Brodrick, 2009; Shapira & Mongrain, 2010) of four partial replications of the Seligman et al. (2005) study showed statistically significant increases in participant well-being. They also remarked that the clearest support for the effectiveness of the positive psychology interventions derived from the original study itself.

Similarly equivocal results are to be found among studies that do not use web-based interventions. For example, in a review of PPIs, Bolier et al. (2013) found that although there are some randomized controlled trials that have shown large, positive results (e.g., Buchanan & Bardi, 2010; Emmons & McCullough, 2003; Quoidbach, Wood, & Hansenne, 2009), among those using a placebo control or alternative-treatment control, effect sizes have generally been small. Moreover, higher quality studies tended to show smaller effect sizes than lower quality studies, and a majority of studies included an effect-size estimate of zero within the 95% confidence interval.

Please address correspondence to: Mark R. Diamond, School of Psychology, University of Tasmania, Private Bag 30, Hobart TAS 7001, Australia

Published online in Wiley Online Library (wileyonlinelibrary.com/journal/jclp). DOI: 10.1002/jclp.22328

More recently, Mongrain and Anselmo-Matthews (2012) reported the results of a large-scale Internet-based partial replication of Seligman et al. (2005). In their study Mongrain and Anselmo-Matthews (2012) sought to match the demand characteristics of two placebo control conditions with that of the PPIs; they did this by including in their participant instructions a more convincing rationale for why the placebo exercise might improve well-being than was provided in the original study. Having found no difference between the effect of the PPIs and their two control groups (enhanced early memories and early positive memories) Mongrain and Anselmo-Matthews (2012) concluded that the effect of all the interventions, including both the PPIs and the placebos, was a consequence of the activation of positive, self-relevant information in participants rather than powerful specific ingredients in the exercises as had previously been suggested (Seligman et al., 2005).

Although the conclusion of Mongrain and Anselmo-Matthews (2012) is clearly supported by their results, the design of their replication trial was predicated on the assumption that the differential effect found by Seligman et al. (2005) between PPIs and a placebo that did not have high demand characteristics was itself replicable. But is it? And what is at stake if it is not? Are PPIs, when used in a web-based context, able reliably to produce important psychological effects? Assuming that PPIs are powerful, do they contain therapeutic components that meet the twin criteria of (a) being distinct from factors common to low-demand, theoretically neutral therapeutic interventions such as unenhanced placebo conditions, and (b) being distinguishable from other therapeutic approaches (e.g., cognitive therapy) by virtue of containing components that are uniquely and explicitly developed within the positive psychology framework?

The difficulty of answering these questions is compounded by two factors. No published reexamination of Seligman et al.'s study has actually compared the original PPIs with the original low-demand, theoretically neutral placebo control. A further complicating factor is that the choice of outcome measures has not been consistent across the various studies, so it is not obvious that the psychological variables said to be influenced in later studies are the same as those which Seligman et al. sought to change (see Table 1).

Web-Based Delivery for Public Health

Whether or not the outcomes of web-based PPIs are the product of potent, novel therapeutic ingredients is important because web-based interventions, where they are effective, offer considerable advantages over traditional approaches in the domain of public health. Because web-based interventions can frequently be delivered anonymously, at low cost, and to very large populations, both their cost-effectiveness and their audience reach have the potential to greatly exceed that of traditional delivery formats (Teachman, 2014) and to result in superior effectiveness and efficiency overall. Despite these attractive aspects, if the only active components of PPIs are nonspecific treatment ingredients, then research effort might be better directed toward how to harness these active but generic components in web-based deliveries of public health interventions rather than being directed towards further investigation of PPIs. As Teachman (2014) noted, a critical research issue in the area of web-based psychological interventions is to discover the key mechanisms of change—without this knowledge it will not be possible to enhance treatment effects and "distill the necessary ingredients" (p. 86) for therapeutic change.

The possibility that the effects found by Seligman et al. (2005) might be attributable to nonspecific effects is heightened by the milieu in which the study took place. Participants in Seligman et al.'s (2005) study were recruited via the website www.authentichappiness.org, a site that is mentioned frequently in the book *Authentic Happiness* (Seligman, 2002), that was, at the time of recruitment for the study, on several best-seller lists. During the recruitment period, the website was attracting an average of 300 new registrants each day (Seligman et al., 2005, p. 415). It is possible that these new registrants were weighted toward people who had been influenced by *Authentic Happiness*—either by reading the book or by hearing about it—and who had an enhanced interest in, and expectation of, positive psychology interventions. Thus, the effects found by Seligman et al. (2005) might be due to participant expectations or indirect therapist (experimenter) effects (Klein et al., 2012).

Table 1
Web-Based Studies of the Effectiveness of Positive Psychology Interventions in Altering Psychological Variables

Authors	Design	Intervention	Control group	Outcome measures
Seligman et al. (2005)	RCT	Three good things; gratitude visit; using signature strengths; identifying strengths; you at your best	Placebo (Writing about early memories)	SHI, CES-D
Goldstein (2007)		Cultivating sacred moments	Placebo (writing about daily activities)	SWLS, RPWB
Abbott, Klein, Hamilton, & Rosenthal (2009)	RCT	"Resilience online"	Waiting list control	WHOQOL-BREF, AHI, DASS-21
Mitchell et al. (2009)	RCT	Using signature strengths; problem solving	Placebo (information only)	SWLS, PANAS, PWI-A, OTH, DASS-21
Shapira & Mongrain (2010)	RCT	Self-compassion; optimism	Placebo (Writing about early memories)	SHI, CES-D
Mongrain & Anselmo-Matthews (2012)	RCT	Three good things; using signature strengths	Placebo(s)	SHI, CES-D
Schueller & Parks (2012)	RCT	Active constructive responding; gratitude visit; life summary; three good things;Savoring; strengths)	No intervention (assessment only)	CES-D

Note. Adapted with additions from Mitchell et al. (2010) and Bolier et al. (2013).
AHI = Authentic Happiness Inventory; CES-D = Center for Epidemiological Studies, Depression Scale; DASS-21 = Depression Anxiety Stress Scale, 21 item version; OTH = Orientations to Happiness; PANAS = Positive and Negative Affect Scale; PWI-A = Personal Wellbeing Index – Adult version; RCT = randomized trial with self-selection but random allocation to group; RPWB = Ryff's Scales of Psychological Well-being; SHI=Steen Happiness Index; SWSL = Satisfaction With Life Scale; WHOQOL-BREF = World Health Organization Quality of Life, Brief version.

Our aims in replicating the study by Seligman et al. (2005) were twofold: first, to compare in a randomized control design the three strongest PPIs used by Seligman et al. with the same theoretically-neutral placebo control that they used; second, to replicate the web-based delivery of PPIs used by Seligman et al. (2005) while seeking to minimize participant expectation effects related to the book *Authentic Happiness*.

Method

Participants

Recruitment. Information about the existence of the study was made available in a media release that was picked up by several Australian newspapers, a local radio station, and an online campus news service. The media release referred interested people to a website (www.happiness-study.org) where they could enroll in the study if they wished.

Characteristics. The initial sample comprised 295 participants aged 18–83 years (mean 43 years), the majority (85%) of whom were female. Most participants (75%) had either a bachelor- or postgraduate-level tertiary qualification and most (76%) described their income as being "average or above."

The mean level of depressive symptomatology among the initial sample (Center for Epidemiologic Studies Depression Scale [CES-D; Radloff, 1977] measure of depression, $M = 15.65$, $SD = 10.52$) was below the threshold for mild depression (threshold = 16; Geisser, Roth, & Robinson, 1997).

Remuneration. No direct financial incentive was offered to participants, but they were informed that if they completed all the follow-up questionnaires, then they would be entered into a lottery draw to win a prize. The offered prize was a single voucher worth AU$100 redeemable at Amazon.com. The choice of a voucher rather than a monetary prize reflected the preference of the university research ethics committee for the use of nonmonetary incentives to encourage participation.

Measures

Authentic Happiness Inventory (AHI). The AHI (Park, Park, & Peterson, 2010) is the name given to a revision and slight extension of the earlier Steen Happiness Index (SHI; Seligman et al., 2005). The SHI was created in response to the developers' belief that previous measures of happiness did not provide good discrimination within the higher range of happiness scores. The AHI is a 24-item multiple-choice measure. For each item, respondents must pick from among five statements the one that best describes them at the present time. Responses are scored on a 5-point scale ranging from 1 (*negative*) to 5 (*extreme positive*). Scores on the items are summed; higher scores indicate higher levels of happiness. The minimum and maximum possible scores are 24 and 120, respectively. Shepherd, Oliver, and Schofield (2015) report that the internal consistency of the AHI is high (Cronbach's alpha = .92), as is the test-retest reliability of the AHI as measured using an intraclass correlation coefficient model (ICC = .92).

CES-D. The CES-D (Radloff, 1977) is a widely used 20-item self-report measure of depressive symptomatology. Respondents rate the frequency with which they have experienced each symptom over the prior week using a 4-point scale ranging from 0 (*rarely or none of the time*) to 3 (*most or all of the time*). Scores on the items are summed; higher scores indicate higher levels of depression. The CES-D has been shown, in a broad range of populations, to have good psychometric properties including concurrent validity, reliability, and consistency (Corcoran & Fischer, 1987; Miller, Anton, & Townson, 2008), with Cronbach's alpha in the range .85–.90 (Radloff, 1977; Santor, Zuroff, Ramsay, Cervantes, & Palacios, 1995).

Procedure

The design of the study paralleled that used by Seligman et al. (2005), including an intervention period lasting just 1 week. The primary difference between this study and that of Seligman et al. (2005) was that we compared the effects of the placebo control with only the three strongest of the original five PPIs.

Potential participants who visited the website at www.happiness-study.org were able to enroll in the study through the site. After giving their informed consent, which included consent to have their (de-identified) data placed in an open-source data repository, participants responded to a series of questions eliciting contact (e-mail) and demographic information, after which they completed the AHI and CES-D measures. Once having completed the two measures, participants were informed that they would receive an e-mail containing instructions about the exercise they were to complete in the succeeding week.

Each participant was randomly assigned to one of four conditions and e-mailed directions as follows.

Gratitude visit. Participants were instructed to write and deliver in person a letter of appreciation to someone who had been kind to them but whom they had never properly thanked.

Three good things in life. Participants were instructed that for a week, they were to write down three good things that happened each day, together with a causal explanation for each thing.

Using signature strengths. Participants were e-mailed a special link to a page on the study website (hidden from participants who were not in the using signature strengths condition) where they completed a questionnaire that gave them feedback about their top five character strengths (Diamond, O'Brien-Malone, & Woodworth, 2010; Peterson, Park, & Seligman, 2005a). Participants were instructed to use one of their five signature strengths in a new and different way for a week.

Early memories: placebo control. Participants were instructed to write about their early memories each night for a week.

Participants in all conditions were encouraged to print out, or write down, the instructions for their activity and keep the instructions accessible in the weeks to come.

The timing and content of all e-mails to participants was managed using an automated e-mailer (Coleman, 2010). The e-mailer was used throughout the study to send scheduled follow-up reminders to participants, including a reminder e-mail in the middle of the week assigned for the intervention activity, in which the instructions for the assigned activity were repeated.

In addition to completing the AHI and CES-D at pretest—that is, immediately before the weeklong intervention period—participants were asked to complete the AHI and CES-D on five further occasions: (a) posttest, immediately having done the prescribed exercises for one week; (b) 1 week after the posttest; (c) 1 month after the posttest; (d) 3 months after the posttest; and (e) 6 months after the posttest. The six measurement occasions are referred to in the remainder of this paper as: pretest, posttest, 1-week follow-up, 1-month follow-up, 3-month follow-up, and 6-month follow-up.

Results

Attrition

A total of 295 participants enrolled in the study, completed the initial questionnaires and were assigned to either a PPI or the control condition. Of those, 147 (49.8%) completed at least the posttest and 72 (24.4%) completed all the requirements of the study—a completion rate almost identical with that reported by Mongrain and Anselmo-Matthews (2012). The full pattern of attrition and measurement completion is shown in Figure 1.

There was no difference between participants who failed to complete one or more assessments (dropouts) and those who completed all assessments (non-dropouts) in terms of sex, $\chi^2(1, N = 295) = 0.01$, $p = .92$; intervention group, $\chi^2(3, N = 295) = 6.15$, $p = .11$; education (Wilcoxon $W = 8154$, $p = .83$); income (Wilcoxon $W = 8552$, $p = .37$); initial happiness scores, $t(293) = 0.26$, $p = .79$; or initial depression scores $t(293) = 0.44$, $p = .66$. Dropouts were on average 5.4 years younger than non-dropouts ($M = 42.5$ vs. 47.8), Welch $t(126.1) = 3.28$, $p = .001$.

Power. In addition to its implications for the general interpretation of the study results, the attrition rate also has implications for the power of the study to detect meaningful changes in participants' happiness and depression. We followed the approach to power analysis recommended by O'Keefe (2007; see also Hoenig & Heisey, 2001) to estimate the power that our study would have in detecting the modest intervention-by-time interaction effects found by Mongrain and Anselmo-Matthews (2012). Using the GPower software (Faul, Erdfelder, Lang, & Buchner, 2007) with $N = 72$, $\alpha = .05$, Cohen's $f = 0.166$, and the actual correlation between repeated measures in our data (minimum $r = .63$), we find that power ($1-\beta$) is .973; with $\alpha = .01$, power still exceeds .90.

Replication

Seligman et al. (2005), and later Mongrain and Anselmo-Matthews (2012), used mixed-design analyses of variance (ANOVAs) as the primary tool for determining whether the happiness and depression of participants had been differentially affected by PPIs, excluding from their analyses those participants with incomplete data. For the purpose of direct comparison and replication, we conducted similar analyses, including in the ANOVAs only those participants who completed all the study requirements ($N = 72$). In the next section, we report additional analyses using all enrolled participants.

Table 2 shows means and standard deviations of happiness (AHI) and depression (CES-D) measures for each intervention group on each of the six measurement occasions (pretest, posttest, 1-week follow-up, 1-month follow-up, 3-month follow-up, 6-month follow-up).

Happiness (AHI). A 4 × 6 (Intervention × Occasion) mixed-design ANOVA was conducted with happiness (AHI) scores as the dependent measure. Although there was a significant main effect for Occasion, $F(5, 340) = 5.74$, $p < .001$, $\eta_p^2 = .078$, indicating that the happiness scores of participants generally increased over the course of the study, the Intervention × Occasion interaction was nonsignificant, $F(15, 340) = 0.499$, $p = .94$, $\eta_p^2 = .020$, which suggests that there was no greater effect of the PPIs relative to the placebo on changes in AHI score over time.

Depression (CES-D). Scores of depressive symptomatology (CES-D) were analyzed using the same 4 × 6 (Intervention × Occasion) mixed-design ANOVA as for the happiness scores. The significant effect for Occasion indicates that depressive symptomatology among participants changed over the course of the study, $F(5, 340) = 2.72$, $p = .02$, $\eta_p^2 = .038$. But as was the case with the happiness scores, the Intervention × Occasion interaction was nonsignificant,

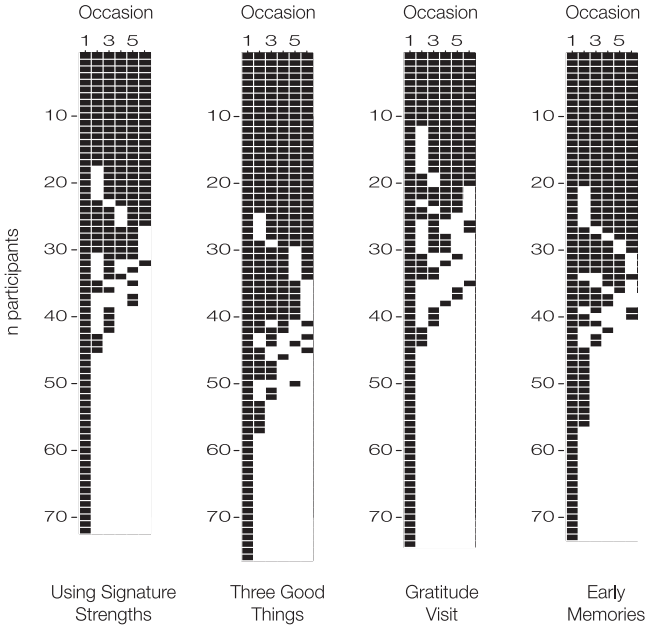

Figure 1. Participant inclusion and attrition for each of the four intervention groups as a function of measurement occasion.
Note. Within intervention groups, rows represent individual participants. Solid black rectangles indicate completion of both the AHI and CES-D on a particular measurement occasion by the relevant participant. White spaces indicate noncompletion of the AHI and of the CES-D. Participants are ordered from top to bottom by number of completed measurements.

Table 2
Means and Standard Deviations (Shown in Parentheses) for Happiness (AHI) and Depression (CES-D) by Intervention and Occasion

Measure and intervention	Pretest	Posttest	1-week follow-up	1-month follow-up	3-month follow-up	6-month follow-up
Happiness (AHI)						
Using signature strengths (n = 17)	67.94	72.65	71.82	74.59	75.94	75.41
	(12.41)	(15.02)	(14.94)	(14.49)	(15.22)	(13.40)
Three good things (n = 24)	69.96	72.58	73.71	72.08	74.25	74.08
	(11.84)	(10.89)	(10.92)	(13.08)	(14.17)	(15.27)
Gratitude visit (n = 11)	72.55	75.27	74.82	75.91	75.18	75.91
	(14.51)	(14.67)	(14.56)	(14.34)	(15.78)	(14.97)
Early memories (n = 20)	71.60	74.20	74.30	75.40	75.65	74.55
	(11.94)	(11.66)	(11.43)	(12.30)	(12.82)	(12.66)
Depression (CES-D)						
Using signature strengths (n = 17)	14.00	13.76	14.18	11.94	10.76	13.00
	(12.17)	(16.26)	(16.30)	(15.87)	(12.36)	(14.82)
Three good things (n = 24)	14.46	12.13	10.92	12.42	13.25	15.13
	(8.71)	(9.35)	(8.94)	(10.36)	(10.63)	(15.49)
Gratitude Visit (n = 11)	14.18	11.00	12.09	10.36	11.00	12.09
	(13.27)	(9.56)	(12.55)	(10.89)	(14.96)	(14.89)
Early memories (n = 20)	11.95	7.40	9.70	7.55	8.50	10.65
	(8.19)	(6.97)	(9.40)	(8.07)	(7.44)	(11.97)

Note. Only participants who completed all the study requirements (N = 72) and whose data were included in the mixed-design ANOVAs are included here. AHI = Authentic Happiness Index (Park et al., 2010); CES-D = Center for Epidemiologic Studies Depression Scale (Radloff, 1977).

$F(15, 340) = 0.622$, $p = .86$, $\eta_p^2 = .026$, suggesting that there was no differential decrease in CES-D scores in the PPI groups over that found in the placebo group.

Extension

The power of the mixed-design analyses of variance to detect differential changes in participants in the various intervention conditions was limited by the fact that only one in four of the initial participants completed all five follow-up questionnaires. To overcome those limitations, we conducted additional analyses (see, for example, Cleveland, 1993), including graphical presentations of the data followed by multilevel modeling.

Figures 2 and 3 show the mean AHI score and CES-D score for each of the four intervention groups at six points in time (corresponding to six measurement occasions). However, so as to give an accurate indication of the rate of change of the dependent variable, the measurement points are shown on the horizontal axis using a continuous time metric, rather than equally spaced categorical "occasions." It is apparent that from the pretest to the 1-week follow-up, there is a rapid rise in AHI scores (Figure 2) and a correspondingly rapid decrease in CES-D scores (Figure 3). A smaller change is manifest over the following 3 weeks—that is, to the 1-month follow-up—for three of the four intervention groups, and thereafter there is only very gradual change in the mean score of any of the intervention groups. Visual inspection of the 95% confidence intervals lends support to the conclusions from the mixed-design analyses of variance: The intervention groups appear indistinguishable from each other and from the placebo group in their pattern of change over time on each of the two outcome variables.

Multilevel modeling. We reexamined participants' AHI and CES-D data using an intention-to-treat approach (Newell, 1992) based on intervention assignment ($N = 295$). Multilevel linear mixed-effects models, with maximum-likelihood estimation, were used to relate longitudinal changes in AHI and CES-D scores to intervention group.

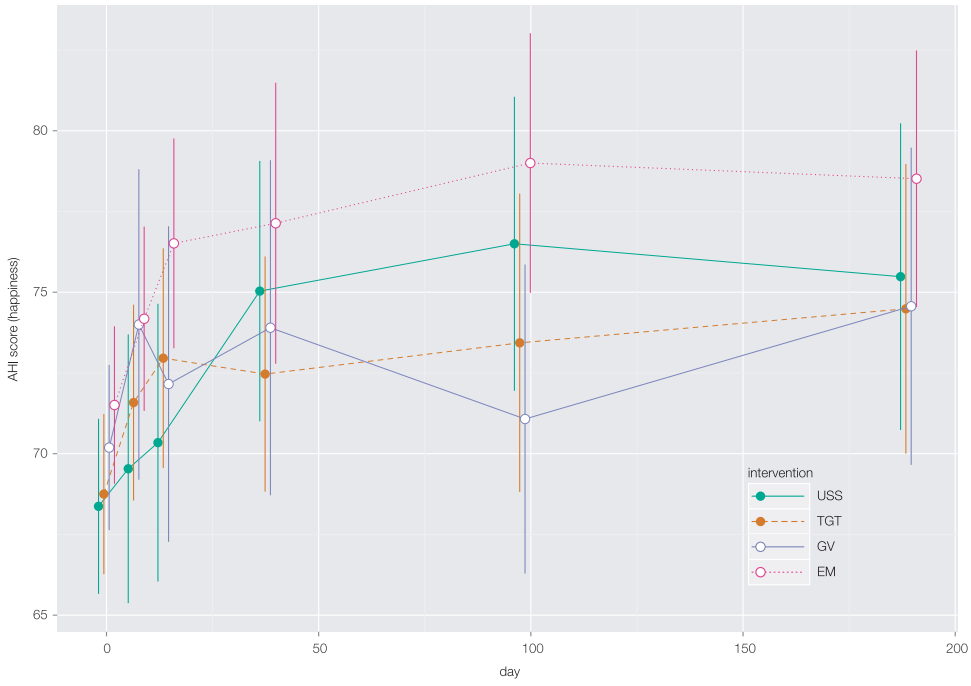

Figure 2. Mean happiness (AHI score) for each of the four intervention groups as a function of time. Note. USS: solid green circles, solid green line; TGT: solid orange circles, dashed orange line; GV: open gray circles, solid gray line; EM: open red circles, dotted red line. Vertical bars show 95% confidence intervals. Group means are offset horizontally for clarity.

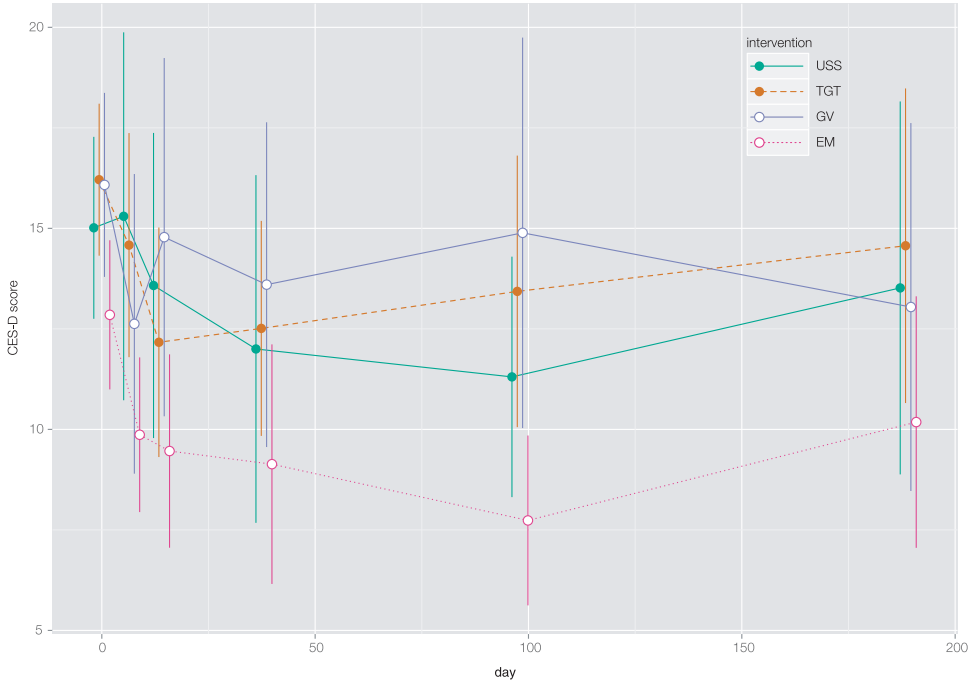

Figure 3. Mean depression (CES-D score) for each of the four intervention groups as a function of time. Note. USS: solid green circles, solid green line; TGT: solid orange circles, dashed orange line; GV: open gray circles, solid gray line; EM: open red circles, dotted red line. Vertical bars show 95% confidence intervals. Group means are offset horizontally for clarity.

Multilevel modeling offers several advantages over traditional mixed-design analysis of variance: In addition to being able to manage unbalanced designs with missing data points, multilevel modeling allows time to be handled easily as a continuous variable. Hence, instead of using Occasion as a categorical within-subject factor in the models, we used information about the actual time that had elapsed (calculated in fractional weeks) between a participant's enrolment in the study and when they completed each of the follow-up measures on the study website.

It is clear from Figures 2 and 3 that the within-intervention group trajectories of both AHI and CES-D scores are not linear over the 6 months of the study; rather, they are approximately linear over the initial period from the pretest to 1-week follow-up and (considered separately) over the period from the 1-month to 6-month follow-up. Consequently, we fitted piecewise linear splines (Harrell, 2015) to the AHI and CES-D data, using maximal models with random intercepts and random slopes (Barr, Levy, Scheepers, & Tily, 2013) and a single spline knot midway between the 1-week follow-up and 1-month follow-up. In practice, we fitted the piecewise splines by coding two separate time-related variables: One variable (TIME1) measured the overall time that elapsed between a participant completing the pretest and each subsequent measure; the other variable (TIME2) measured the time between the spline knot and a participant completing each later measure (Singer & Willett, 2003). The first level of the two-level model may be written as:

$$Y_{ij} = \pi_{0i} + \pi_{1i} TIME1_{ij} + \pi_{2i} TIME2_{ij} + \varepsilon_{ij},$$

where, Y_{ij}. is the response of participant i at time j, π_{0i} represents the intercept of the change trajectory for participant i, π_{1i} represents the slope of the trajectory of change for participant i, π_{2i} represents the change of slope following the spline-knot, and ε_{ij} represents random measurement error.

Level 2 of the model is formulated as:

$$\pi_{0i} = \gamma_{00} + \gamma_{01} INTERVENTION_i + \zeta_{0i},$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11} INTERVENTION_i + \zeta_{1i},$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21} INTERVENTION_i + \zeta_{2i},$$

with

$$\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right)$$

and

$$\begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} \\ \sigma_{02} & \sigma_{12} & \sigma_2^2 \end{bmatrix} \right).$$

The model can be read as indicating that the intercept, π_{0i}, and slopes, π_{1i} and π_{2i}, for each participant i are level 2 outcomes associated with the predictor *INTERVENTION* (i.e., the Intervention factor). In addition, each of the level 2 components is associated with a separate residual—ζ_{0i},ζ_{1i} and ζ_{2i}—that allows the level 1 parameters $\pi_{.i}$ for each participant to differ randomly from those of other participants. It is an empirical matter whether the off-diagonal elements of the covariance matrix are nonzero.

For readers more familiar with the composite specification of multilevel models, the model collapses to:

$$\begin{aligned} Y_{ij} = {} & \gamma_{00} + \gamma_{01} INTERVENTION_i + \gamma_{10} TIME1_{ij} + \gamma_{11} INTERVENTION_i TIME1_{ij} \\ & + \gamma_{20} TIME_{ij} + \gamma_{21} INTERVENTION_i TIME2_{ij} + TIME1_{ij}\zeta_{1i} + TIME2_{ij}\zeta_{2i} + \zeta_{0i} + \varepsilon_{ij} \end{aligned}$$

In terms of examining the differential effects of the interventions on happiness and depression,

Table 3
Parameter Estimates for Fixed Effects in the Linear Mixed-Effects Models for Happiness (AHI) and Depression (CES-D)

Outcome	Fixed effect	Estimate	*SE*	*t*	*df*[a]	*p*[b]
Happiness (AHI)	*Intervention* (γ_{01})					
	USS	0[c]				
	TGT	0.61	2.13	0.28	315.7	.78
	GV	1.78	2.15	0.83	319.7	.41
	EM	3.28	2.15	1.52	316.1	.13
	Time					
	TIME1[d] (γ_{10})	1.34	0.41	3.26	186.2	< .01
	TIME2[d,e] (γ_{20})	−1.24	0.46	2.70	179.5	< .01
	Interaction Intervention × *TIME1*[d] (γ_{11})					
	USS	0[c]				
	TGT	−0.32	0.55	0.59	185.7	.56
	GV	−0.36	0.59	0.60	198.7	.55
	EM	−0.23	0.56	0.41	192.5	.68
	Interaction Intervention × *TIME2*[d,e] (γ_{21})					
	USS	0[c]				
	TGT	0.24	0.61	0.39	179.8	.70
	GV	0.25	0.67	0.37	193.9	.71
	EM	0.18	0.63	0.29	185.8	.77
Depression (CES-D)	*Intervention* (γ_{01})					
	USS	0[c]				
	TGT	0.91	1.77	0.52	292.1	0.61
	GV	0.91	1.79	0.51	298.4	0.61
	EM	−2.63	1.78	1.47	292.6	0.14
	Time					
	TIME1[d] (γ_{10})	−0.87	0.37	2.36	172.1	< .05
	TIME2[d,c] (γ_{20})	0.88	0.42	2.10	164.9	< .05
	Interaction Intervention × *TIME1*[d] (γ_{11})					
	USS	0[c]				
	TGT	−0.22	0.49	0.45	173.3	.65
	GV	0.00	0.53	0.00	193.5	.99
	EM	−0.12	0.50	0.23	182.2	.82
	Interaction Intervention × *TIME2*[d,e] (γ_{21})					
	USS	0[c]				
	TGT	0.36	0.56	0.65	167.2	.51
	GV	−0.01	0.61	0.02	186.9	.98
	EM	0.17	0.58	0.30	174.8	.77

Note. SE = standard error; USS = using signature strengths; TGT = three good things in life; GV = gratitude visit; EM = early memories (placebo control).
[a]Welch–Satterthwaite pooled degrees of freedom.
[b]*p* values should be treated with caution (Bates et al., 2015).
[c]Estimate is zero because this intervention group is used as the reference (baseline) group for the parameter estimates of the remaining groups.
[d]Parameter estimates for TIME1 and TIME2 relate to a time-metric in units of weeks.
[e]Parameter estimates involving TIME2 indicate the *change of slope* from TIME 1, not absolute slope.

the parameters in the fitted model of greatest interest were those relating to the fixed-effect interaction between the intervention group and time (parameters γ_{11} and γ_{21}). The parameters relating to the main effects for time (parameters γ_{10} and γ_{20}) are of interest insofar as they describe the average trajectory of change of participants over the course of the study. Parameter estimates, obtained using R (R Core Team, 2015) with the *lmer* package (Bates, Mächler, Bolker, & Walker, 2015), are shown in Table 3.

The results from the multilevel models mirror those of the mixed-design ANOVAs reported in the previous section. AHI and CES-D scores are shown to change over time, and the slope of the trajectory is different in the later epoch from that in the early epoch. Critically, however, no significant effect was found in either the AHI model or the CES-D model for the interaction between intervention group and time in either of the two epochs (parameters γ_{11} and γ_{21}), indicating that the four treatment interventions cannot be distinguished in their effect on either happiness levels or depressive symptomatology. On average, the AHI score of participants increased by 1.34 per week over the period from pretest to just after the 1-week follow-up, and similarly the CES-D declined by an average of 0.87 per week. Interestingly, in the case of both AHI and CES-D, the parameter for TIME2 is approximately the additive inverse of that for TIME1 (i.e., $\gamma_{10} + \gamma_{20} \cong 0$), leading to the conclusion that little change occurred in the mean level of participants' happiness or depression after the 1-week follow-up.

Discussion

There were two goals of the present study: first, to compare in a randomized control design the three strongest PPIs used by Seligman and colleagues (2005) with the same theoretically neutral placebo control that they used; second, to replicate the web-based delivery of PPIs used by Seligman and colleagues (2005) while seeking to minimize participant expectation effects.

Participants in the present study engaged in PPIs identical to those used by Seligman et al. (2005) and over the course of 6 months showed improvements in both happiness and depression. A more fine-grained analysis of AHI and CES-D scores using multilevel modeling suggested that most of the change in levels of happiness and depression occurred during the first two weeks of the study—from pretest to the 1-week follow-up—with little change occurring during the remainder of the study period. Although our methodology closely matched that used by Seligman and colleagues (2005) and Mongrain and Anselmo-Matthews (2012), the changes that we observed in happiness and depression were less exceptional than in either of those studies. Further, in contrast to the findings reported by both Seligman and colleagues (2005) and Mongrain and Anselmo-Matthews (2012), we found that participants in the PPI groups showed no greater increase in happiness, nor did they show any greater reduction in depression, than did participants who engaged in the placebo control.

Given this pattern of results, the finding to be explained is not why the happiness and depression of participants in the PPI groups failed to improve. They did improve. Rather, the result that requires explanation, and that distinguishes the outcome of this study from that of Seligman and colleagues (2005) and Mongrain and Anselmo-Matthews (2012), is why participants in the placebo control condition in this study showed similar sustained improvements to those in the PPI groups. An alternative approach to the same issue is to ask why, although participants in the PPI groups did improve, they did not show enhanced effects relative to the placebo-controlled group. In the sections that follow, we consider several possibilities.

Differing Participant Expectations of Efficaciousness

If participants in the placebo-controlled condition (writing about early memories) believed that the intervention was a powerful PPI, then one might anticipate that its positive effects would be increased relative the effects that would be observed in the absence of such expectations (Critelli & Neumann, 1984; Schmidt, Braun, Wager, & Shohamy, 2014). Conversely, if participants in the three PPI groups had low expectations of the effect of the PPIs, then one might expect their observed effectiveness to be reduced.

However, militating against group-dependent expectations as being an explanation for our results is the fact that, with the exception of a description of the intervention activity in which each participant was to engage, we used a standardized set of instructions for all participants. Moreover, following the protocol of the original study by Seligman and colleagues, the instructions, including the description of the intervention activity, were delivered by e-mail or via the study website, rather than in person, as were all follow-up reminders and the AHI and CES-D.

Accordingly, there appears to have been little opportunity for differences in group expectations to have been created by the study procedure.

Context Effects

There have been numerous examples in the psychological, educational, and health literature of interventions that have been highly successful in one place or at one time and failed to produce similar effects in another place or at another time despite having been implemented according to the original successful protocol (Bohrstdet & Stecher, 2002; Cartwright & Hardie, 2012).

Broadly, the reasons for the later failures of interventions that have proved their worth in randomized control trials appear to be subtle differences in the cultural and temporal context of the second place or the later time. That possibility is particularly germane to the present study. Our failure to find significant intervention-dependent effects of the PPIs versus the placebo-controlled on the outcome measures might result from cultural differences between Australians and North Americans that, in an Australian population, diminish the effect of PPIs that have their roots in culturally discordant notions of happiness (for a discussion of cultural differences in perceptions of, and desire for, happiness, see, for example, Blanchflower & Oswald, 2005; Christopher & Hickinbottom, 2008; Joshanloo & Weijers, 2014; Oishi, Graham, Kesebir, & Galinha, 2013).

If that is the case, then the improvements that we observed in participants in the PPI groups are likely to have been brought about by the same kinds of nonspecific ingredients as might account for the improvements observed in the placebo-controlled group. If that explanation is correct, then the pattern of results we observed should be replicable with an Australian sample, whereas the pattern of results found by Seligman and colleagues should replicate using an American sample. This would provide an interesting focus for further research.

Similarly, if the impressive findings reported by Seligman and colleagues (2005) were partly mediated by having been conducted within the milieu created by the launch and popularity of the book *Authentic Happiness*, then one might anticipate the effects to be diminished in a different context. Although there is no direct evidence to support that explanation, Mongrain and Anselmo-Matthews (2012) reported that their study "yielded results that were more modest" (p. 387) than those of Seligman and colleagues, as indeed are our results.

Selection Bias

Parks, Della Porta, Pierce, Zilca, and Lyubomirsky (2012) reported having found two distinct categories of online happiness seekers: nondistressed happiness seekers who did not suffer from clinical depression, and distressed happiness seekers who did. The authors suggested that because happiness seekers do not form a single, homogeneous group, individuals might differ in their responses to PPIs depending on the symptoms they experience. In particular, individuals with depression might find it difficult to perform activities that entail engagement (see, for example, Peterson, Park, & Seligman, 2005b), such as those required by the three PPIs.

If participants who enrolled in our study had systematically higher levels of depression than participants in the studies of Mongrain and Anselmo-Matthews (2012) and Seligman and colleagues (2005), then the PPIs might have had limited effectiveness. However, the available evidence does not appear to support that possibility. Whereas the mean CES-D score of participants in the study by Mongrain and Anselmo-Matthews (2012) was 20.4, in this study the mean CES-D score of participants (14.2) was identical with that which can be estimated from the graphical data presentations in the study by Seligman and colleagues (2005).

Conclusion

Although our results do not support the conclusion that the three PPIs we examined were more therapeutically active than a low-demand therapeutically generic placebo, they do support the view that the PPIs contain ingredients that can bring about significant changes in happiness and depression. From the point of view of a public health psychologist, the apparently equal efficacy

of the interventions used in this study might be of little concern: three novel interventions (PPIs) will have been added to the existing therapeutic armamentarium and all have the desirable attributes described by Teachman (2014) of being able to reach an enormous number of people at a very low cost. Improvements that they effect in the psychological health and well-being of a population could be significant.

From the point of view of a guiding theory for positive psychology, however, the issue is whether the theory that has motivated the construction of the PPIs has actually led to them incorporating specific ingredients that are absent from nominally non-PPIs, and if so, whether those theory-driven ingredients can, or do, lead to enhanced levels of change. Taken together with the finding by Mongrain and Anselmo-Matthews (2012) that an enhanced placebo-controlled condition was as effective as two tested PPIs in improving happiness and reducing depression, our failure to find a difference between the effectiveness of the PPIs and that of the placebo-controlled intervention lends weight to the idea that PPIs might not contain specific ingredients.

Even the most powerful therapeutic interventions rely for their effectiveness on many of the ingredients that they share with other, demonstrably less effective therapies (Greenberg, 2004; Imel & Wampold, 2008), and one of the most important of the shared ingredients appears to be the "client factor" (Tallman & Bohart, 1999). Participants in this study and in the studies by Seligman et al. (2005) and Mongrain and Anselmo-Matthews (2012) share the notable feature of being motivated to improve their happiness and to look in a particular place (the online environment) for a method of achieving that objective and this "client factor" might account for a significant proportion of the shared effectiveness of the interventions. Contextual effects, on the other hand, might work to discriminate between them. A critical research issue then is to discover what interventions work best in what context, when, and with whom.

References

Abbott, J.-A., Klein, B., Hamilton, C., & Rosenthal, A. (2009). The impact of online resilience training for sales managers on wellbeing and work performance. Electronic Journal of Applied Psychology, 5(1), 89–95.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48.

Blanchflower, D. G., & Oswald, A. J. (2005). Happiness and the human development index: The paradox of Australia. National Bureau of Economic Research Working Paper Series, No. 11416.

Bohrstdet, G. W., & Stecher, B. M. (Eds.). (2002). What we have learned about class size reduction in California. Sacremento: California Department of Education.

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), 1–20.

Buchanan, K. E., & Bardi, A. (2010). Acts of kindness and acts of novelty affect life satisfaction. Journal of Social Psychology, 150(3), 235–237.

Cartwright, N., & Hardie, J. (2012). Evidence-Based policy: A practical guide to doing it better. New York: Oxford University Press.

Christopher, J. C., & Hickinbottom, S. (2008). Positive psychology, ethnocentrism, and the disguised ideology of individualism. Theory and Psychology, 18(5), 563–589.

Cleveland, W. S. (1993). Visualizing data. New York: Hobart Press.

Coleman, A. (2010). Infinite responder software (version 148e). Retrieved from http://www.infinite.ibasiscs.downloadable.biz

Corcoran, K., & Fischer, J. (1987). Measures for clinical practice: A source book (Vol. 1). New York: The Free Press.

Critelli, J. W., & Neumann, K. F. (1984). The placebo: Conceptual analysis of a construct in transition. American Psychologist, 39(1), 32–39.

Diamond, M. R., O'Brien-Malone, A., & Woodworth, R. J. (2010). Scoring the VIA survey of character. Psychological Reports, 107(3), 833–836.

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.

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(2), 175–191.

Geisser, M. E., Roth, R. S., & Robinson, M. E. (1997). Assessing depression among persons with chronic pain using the Center for Epidemiological Studies–Depression Scale and the Beck Depression Inventory: A comparative analysis. Clinical Journal of Pain, 13(2), 163–170.

Goldstein, E. D. (2007). Sacred moments: Implications on well-being and stress. Journal of Clinical Psychology, 63(10), 1001–1019.

Greenberg, R. P. (2004). Essential ingredients for successful psychotherapy: Effect of common factors. In M. J. Dewan, B. N. Steenbarger, & R. P. Greenberg (Eds.), The art and science of brief psychotherapies: A practitioner's guide (pp. 231–241). American Psychiatric Publishing: Arlington, VA.

Harrell, F. E., Jr. (2015). Regression modeling strategies: With applicatoins to linear models, logistic and ordinal regression, and survival analysis. New York: Springer International Publishing.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. The American Statistician, 55(1), 19–24.

Imel, Z. E., & Wampold, B. E. (2008). The importance of treatment and the science of common factors in psychotherapy. In Handbook of counseling psychology (4th ed., pp. 249–266). Hoboken, NJ: John Wiley & Sons Inc.

Joshanloo, M., & Weijers, D. (2014). Aversion to happiness across cultures: A review of where and why people are averse to happiness. Journal of Happiness Studies, 15(3), 717–735.

Klein, O., Doyen, S., Leys, C., Magalhães de Saldanha da Gama, P. A., Miller, S., Questienne, L., & Cleeremanis, A. (2012). Low hopes, high expectations: Expectancy effects and the replicability of behavioral experiments. Perspectives on Psychological Science, 7(6), 572–584.

Miller, W. C., Anton, H. A., & Townson, A. F. (2008). Measurement properties of the CESD scale among individuals with spinal cord injury. Spinal Cord, 46(4), 287–292.

Mitchell, J., Stanimirovic, R., Klein, B., & Vella-Brodrick, D. (2009). A randomised controlled trial of a self-guided Internet intervention promoting well-being. Computers in Human Behavior, 25(3), 749–760.

Mitchell, J., Vella-Brodrick, D., & Klein, B. (2010). Positive psychology and the Internet: A mental health opportunity. Electronic Journal of Applied Psychology, 6(2), 30–41.

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.

Newell, D. J. (1992). Intention-to-treat analysis: Implications for quantitative and qualitative research. International Journal of Epidemiology, 21(5), 837–841.

Oishi, S., Graham, J., Kesebir, S., & Galinha, I. C. (2013). Concepts of happiness across time and cultures. Personality and Social Psychology Bulletin, 39(5), 559–577.

O'Keefe, D. J. (2007). Post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: Sorting out appropriate uses of statistical power analyses. Communication Methods and Measures, 1(4), 291–299.

Park, N., Park, M., & Peterson, C. (2010). When is the search for meaning related to life satisfaction? Applied Psychology: Health and Well-Being, 2(1), 1–13.

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.

Peterson, C., Park, N., & Seligman, M. E. (2005a). Assessment of character strengths. In G.P. Koocher, J.C. Norcross & S.S. Hill (Eds.), Psychologists' desk reference (2nd ed.). New York: Oxford University Press.

Peterson, C., Park, N., & Seligman, M. E. (2005b). Orientations to happiness and life satisfaction: The full life versus the empty life. Journal of Happiness Studies, 6(1), 25–41.

Quoidbach, J., Wood, A. M., & Hansenne, M. (2009). Back to the future: The effect of daily practice of mental time travel into the future on happiness and anxiety. Journal of Positive Psychology, 4(5), 349–355.

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.

R Core Team. (2015). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Santor, D. A., Zuroff, D. C., Ramsay, J. O., Cervantes, P., & Palacios, J. (1995). Examining scale discriminability in the BDI and CES-D as a function of depressive severity. Psychological Assessment, 7(2), 131–139.

Schmidt, L., Braun, E. K., Wager, T. D., & Shohamy, D. (2014). Mind matters: Placebo enhances reward learning in Parkinson's disease. Nature Neuroscience, 17(12), 1793–1797.

Schueller, S. M., & Parks, A. C. (2012). Disseminating self-help: Positive psychology exercises in an online trial. Journal of Medical Internet Research, 14(3), 8–18.

Seligman, M. E. (2002). Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment. New York: The Free Press.

Seligman, M. E., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. American Psychologist, 60(5), 410–421.

Shapira, L. B., & Mongrain, M. (2010). The benefits of self-compassion and optimism exercises for individuals vulnerable to depression. Journal of Positive Psychology, 5(5), 377–389.

Shepherd, J., Oliver, M., & Schofield, G. (2015). Convergent validity and test–retest reliability of the authentic happiness inventory in working adults. Social Indicators Research, 124(3), 1049–1058.

Singer, J. D., & Willett, J. B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. New York: Oxford University Press.

Tallman, K., & Bohart, A. C. (1999). The client as a common factor: Clients as self-healers. In M. A. Hubble, B .L. Duncan, & S. D. Miller (Eds.), The heart and soul of change: What works in therapy (pp. 91–131). Washington, DC: American Psychological Association.

Teachman, B. A. (2014). No appointment necessary: Treating mental illness outside the therapist's office. Perspectives on Psychological Science, 9(1), 85–87.