Psychiatry Research 269 (2018) xxx–xxx

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

A randomized controlled trial on a smartphone self-help application (Be Good to Yourself) to reduce depressive symptoms

Thies Lüdtke[a,b,*], Lilian Klara Pult[a,1], Johanna Schröder[a,c], Steffen Moritz[a], Lara Bücker[a]

[a] Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[b] Department of Psychology, UiT – The Arctic University of Norway, Hansine Hansens veg 18, 9019, Tromsø, Norway
[c] Institute for Sex Research and Forensic Psychiatry, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

ARTICLE INFO

Keywords:
e-mental health
Quality of life
Self-esteem
App

ABSTRACT

Depressive symptoms are common, yet only a subgroup of individuals receive adequate treatment. To reduce the treatment gap, several online self-help programs have been developed, yielding small to moderate effects. We developed a smartphone self-help application addressing depressive symptoms. We sought to evaluate its feasibility and efficacy in participants reporting a subjective need for help (a diagnosis of depression was not mandatory). We conducted a randomized controlled trial (N = 90). The primary outcome was a reduction of depressive symptoms measured with the Patient Health Questionnaire-9 (PHQ-9). Secondary outcomes included improved self-esteem (Rosenberg Self-Esteem Scale) and quality of life (WHOQOL-BREF). The intervention group obtained access to the application for four weeks, the wait-list group received access after the post assessment. No group differences emerged in either outcome in intention-to-treat analyses. Per protocol analyses with frequent users (i.e., several times a week or more) yielded a small effect size (η_p^2 = 0.049) at trend level on the reduction of depressive symptoms in favor of the treatment group. However, 39% of the participants did not use the application frequently. Mobile self-help applications represent a promising addition to existing treatments, but it is important to increase patients' motivation to use them.

1. Introduction

Depressive symptoms are common in the general population, with a life-time prevalence of 54.4% for any depressive disorder (Vandeleur et al., 2017), but many individuals remain untreated (Kazdin, 2017). To illustrate, only 56% of the individuals with major depression worldwide receive treatment (Kohn et al., 2004). Possible reasons for this treatment gap are limited capacity of therapists and clinics, geographical restrictions and factors related to the individuals in need of help (e.g., fear of stigmatization, restricted mobility because of physical conditions or apathy; Kohn et al., 2004; Mohr et al., 2010).

The treatment gap exists despite several effective treatments for depressive symptoms. Cognitive behavioral therapy (CBT) is currently the most evidence-based psychotherapy for depression (e.g., Driessen et al., 2016). More recently, methods of the third wave of CBT, for example mindfulness-based therapy (Khoury et al., 2013), and metacognitive therapy (Wells and Papageorgiou, 2004), have proven to be effective in reducing symptoms of depression (Churchill et al., 2013; Normann et al., 2014; Strauss et al., 2014); and also for metacognitive training for depression there are promising findings regarding its efficacy (Jelinek et al., 2015; Jelinek et al., 2016). Traditional psychotherapy, however, may not be capable of handling the anticipated demand for psychosocial interventions in the future (Kazdin and Blase, 2011).

Technology-based interventions, both guided (requiring contact with a therapist) and nonguided (used by patients on their own), are possible complementary or subsidiary approaches to traditional face-to-face interventions. Technology-based interventions could be particularly useful to bridge the gap when patients are untreated since patients must often wait several months for their treatment to start, if treatment is sought or initiated at all. Technology-based interventions are independent of geographic constraints and time (Twomey et al., 2017), they offer anonymity (Young, 2005) and low delivery costs (Warmerdam et al., 2010). Also, the attitude of the general population towards such interventions is positive. According to a national survey (N = 2411), more than one-fourth (26.3%) of Germans could imagine seeking information and help online in case of a mental disorder or emotional distress. Interestingly, willingness to use therapeutic online

* Corresponding author at: Department of Psychology, UiT – The Arctic University of Norway, Hansine Hansens veg 18, 9019, Tromsø, Norway.
E-mail address: t.luedtke@uke.de (T. Lüdtke).
[1] Joint first author.

https://doi.org/10.1016/j.psychres.2018.08.113
Received 12 December 2017; Received in revised form 10 July 2018; Accepted 28 August 2018
Available online 29 August 2018
0165-1781/ © 2018 Elsevier B.V. All rights reserved.

counselling is higher in individuals who frequently use the internet and related devices (Eichenberg et al., 2013). The demand for online counseling will likely increase with growing digitalization in the future. Therefore, technology-based treatments of psychological disorders, also known as e-mental health, have the potential to decrease the treatment gap in the mental health system.

Several psychological online interventions (POIs) for the treatment of depressive symptoms have been developed and evaluated in the last decade, yielding small to medium effects (Cuijpers et al., 2011; Griffiths et al., 2010; Johansson and Andersson, 2012; Karyotaki et al., 2017; Richards and Richardson, 2012). According to a meta-analysis by Karyotaki et al. (2017), unguided POIs show small effects (*Hedges* $g = 0.27$) on depressive symptoms compared to control groups.

While the efficacy of POIs that are accessed through one's private computer is well established, there is a dearth of research on smartphone-based interventions addressing depressive symptoms. So far, however, the evidence is very promising, with an effect size of *Hedges'* $g = 0.38$ in favor of smartphone-based interventions over control conditions (Firth et al., 2017). There are disadvantages associated with smartphone-based interventions, such as the so-called "digital divide", meaning that people without access to the internet cannot benefit from such interventions (Bert et al., 2014), or the limited amount of words that can be displayed on smartphones' relatively small screens. However, these types of interventions also offer manifold advantages. Because smartphones are widely used, one advantage of smartphone-based interventions is that they can reach many people. Studies find that 68% of the population owns a smartphone in developed countries and 37% in developing countries (Poushter, 2016). Furthermore, preliminary evidence indicates that smartphone-based interventions could complement face-to-face therapies (Ly et al., 2015). The combination of face-to-face and technology-based interventions is called "blended therapy". One of the challenges in technology-based interventions is to sustain treatment effects after treatment has ended, and preliminary evidence from obsessive-compulsive disorder research suggests that "booster programs" (i.e., additional therapy sessions after treatment completion) could help to preserve treatment effects (Andersson et al., 2014). Smartphone applications could help to sustain treatment effects by sending frequent reminders via push notifications.

The present study revolves around the efficacy of unguided smartphone self-help applications, as these are widely available and may partially compensate for the dearth of therapists and mental institutions (Eichenberg et al., 2013). Additionally, some researchers recommend that new studies are needed to identify moderators of positive and negative treatment outcomes in e-mental health (Ebert et al., 2016; Schröder et al., 2016), and this applies to smartphone-based interventions as well. One factor that might influence the efficacy of smartphone-based interventions is willingness to change. According to the transtheoretical model by Prochaska and DiClemente (1982), one goes through different stages of change in psychotherapy. Willingness to change problematic behavior was found to be a moderator in a study of transdiagnostic face-to-face treatment (Boswell et al., 2012) as well as in an uncontrolled study of face-to-face treatment for participants with somatoform disorders (Heider et al., 2018). Additionally, the "action"-subscale of willingness to change was found to moderate the effect of an online intervention addressing marijuana use among students (Palfai et al., 2016). Therefore, willingness to change might function as a moderator in smartphone-based treatment as well.

Considering the sparse results to date regarding the efficacy of unguided mobile interventions, the present study aims at expanding the body of research on this form of intervention for depressive symptoms. We developed a CBT-based self-help smartphone application called Be Good to Yourself (German: *Tu Dir Gut*). We hypothesized that using the smartphone application would result in decreased depressive symptoms, increased self-esteem, and a higher quality of life compared to a wait-list control condition. The effects were expected to be moderated by the participants' willingness to change.

2. Methods

The study was approved by the local ethics committee of the University of Hamburg (Germany) and was set up as an online study with random allocation of participants to either the intervention group (smartphone self-help application) or the wait-list control group.

2.1. Participants

Recruitment was carried out via online forums and an outpatient clinic. An invitation to the study was posted in social media groups and internet support networks devoted to depression. These websites and social media groups disseminate information about the disorder and offer a place for people with depression to exchange opinions and advice. The study invitation summarized the basic design as well as the terms and the procedure. It stressed that all participants would receive free access to the self-help application either immediately or after a four-week delay (after the post assessment).

Inclusion criteria were the subjective need for an intervention to reduce depressive symptoms, age 18–65 years, and the possession of an iPhone (the application was available for the iPhone operating system only). When participants reported high suicidal tendencies (measured with the suicide-item on the PHQ-9; Kroenke et al., 2001), they were automatically excluded from the online survey. We informed participants of the reason for their exclusion and provided information on sources of help (e.g., telephone numbers and contact addresses for specialized institutions). Only two participants were excluded for this reason. Recruitment was stopped after 90 participants with valid responses had completed the baseline survey. One participant was excluded afterwards because of age. As for suicidal tendencies, we implemented a "filter" function in the online survey which was meant to exclude participants automatically when they reported to be older than 65. Unfortunately, this automated exclusion did not operate properly so we had to exclude one participant (aged 67) manually afterwards. Another person withdrew informed consent when asked to take part in the post survey. This resulted in a final sample size of $N = 88$ (see Fig. 1 for a flow chart).

2.2. Procedure

A web link in the study invitation directed potential participants to the baseline survey, which was implemented using *EFS Survey* (www.unipark.info). The program prevented multiple logins from the same computer by means of "cookies". On the first page of the baseline survey, the rationale of the study was explained. In the next step, we obtained electronic informed consent from each participant.

At the end of the survey, participants were required to enter their email address to allow the sending of the access codes to those in the treatment group. Those in the wait-list group were notified that they had been allocated to the control group and would receive their access code upon completion of the post survey four weeks later.

Participants were randomly allocated to the treatment or the wait-list group in consecutive order. Allocation was concealed, and no stratification was applied. Within 24 hours, participants in the treatment group received an email with instructions on how to download the self-help application Be Good to Yourself. The participants in the wait-list group were informed that they would receive access to the application after completion of the post survey four weeks later. The rationale for having wait-list control groups in randomized trials was briefly explained.

Four weeks after baseline assessment, we invited all participants via email to take part in the post survey. If necessary, participants were reminded up to three times to participate in the post assessment. Upon entry to the post survey, we requested participants to enter the same email address as for the baseline survey to allow the matching of baseline and post questionnaires. As an incentive for completing the

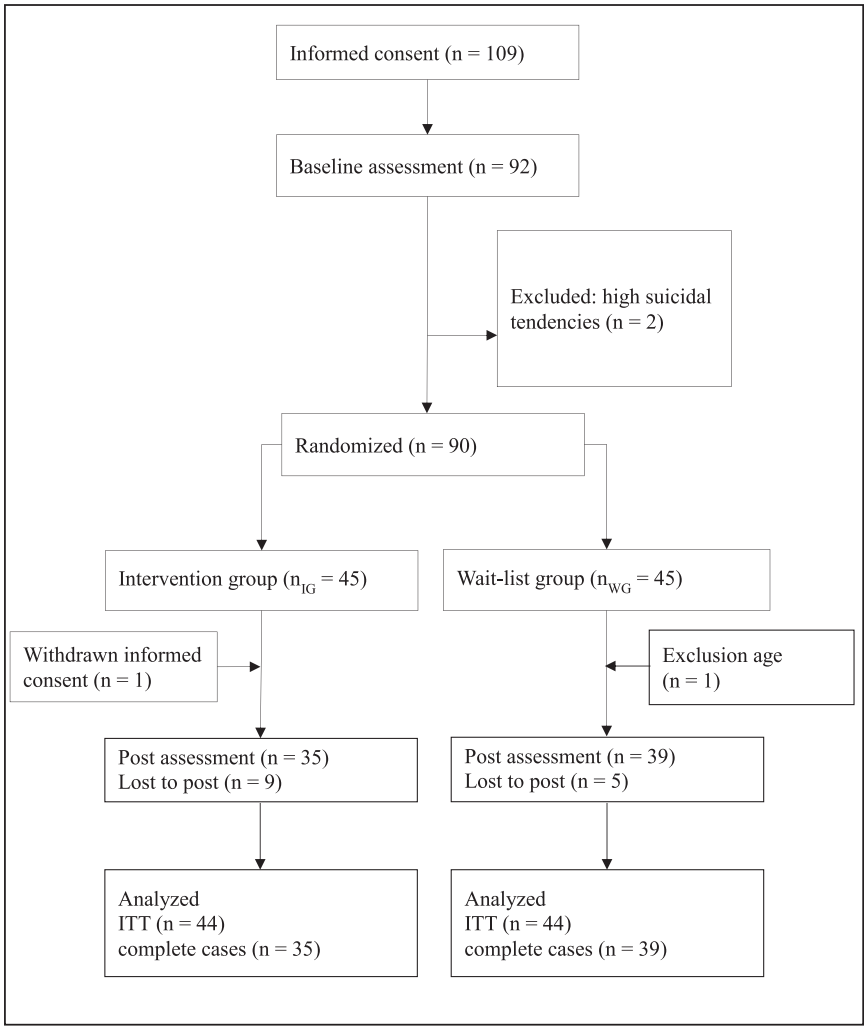

Fig. 1. Flowchart.

post assessment, a PDF file with additional mindfulness-based tasks was available for download at the end of the survey.

2.3. Intervention

The intervention Be Good to Yourself is a newly developed self-help application for the iPhone's operating system iOS (as this was a proof-of-concept study with limited resources, we developed the application for one operating system only) that consists of 40 self-help strategies and exercises. These exercises are based on CBT and its third wave. Exercises are assigned to four categories: cognitive strategies (see Fig. 2), mindfulness-based exercises (see Fig. 3), social-competence skills (see Fig. 4), and activating exercises (see Fig. 5). Be Good to Yourself is intended to serve as a "companion" to the user's daily routine, and it helps users take time for their own psychological well-being. Therefore, each exercise is described in less than 150 words and is easily read and performed in a couple of minutes. Each exercise includes a short psychoeducational section as well as instructions for an exercise.

The exercises appear in fixed order, meaning that all participants who use the intervention receive the same sequence of exercises. However, exercises alternate among the four different categories. Upon opening the application, the title of the exercise is presented. The user can either select it and read the detailed description or skip to the next exercise. A small icon indicates the category of the exercise. On the title screen, the exercise can either be selected or skipped up to three times. When selecting the exercise, a description appears with a countdown that prevents premature closing. This way, the user needs to take at least a couple of minutes to go through the exercise. Following each exercise, a motivational screen appears to reinforce the user's taking time for his or her own psychological well-being (see Fig. 6). The application sends daily reminders to encourage the user to take time for him- or herself. The number of reminders and the times they are sent can be adjusted individually in the settings.

2.4. Assessments

The baseline survey included the following sections: demographic questions (e.g., age, gender, education), medical history (e.g., previous and current treatments, medication, diagnoses, profession of person who diagnosed depression), and a psychological section to assess depressive symptoms, quality of life, self-esteem, and willingness to change. These sections are described in greater detail below. The post survey contained the same questionnaires as the baseline survey (see Sections 2.4.1–2.4.4). In addition, completers of the treatment condition were asked how often they had used the application (not at all, once in four weeks, weekly, several times a week, daily, several times a day). If participants affirmed having used the intervention, we asked them to evaluate how effective, comprehensive, applicable, and appealing they perceived it to be (see Section 2.4.5). Participants were also given the option to leave comments in a text box.

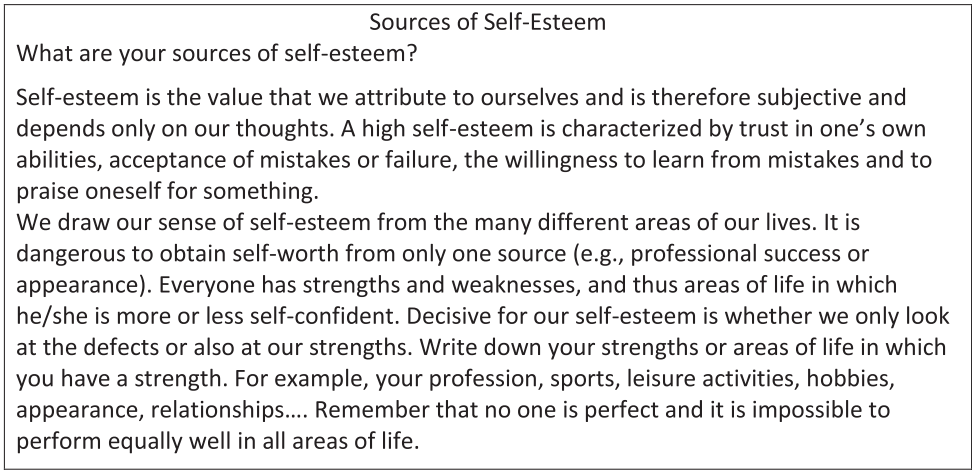

Fig. 2. Example of a task in the category cognitive strategies.

2.4.1. Patient Health Questionnaire (PHQ-9)

The self-report questionnaire PHQ-9 served as the primary outcome (Kroenke et al., 2001). We used the German version by Löwe et al. (2002). The PHQ-9 is an extract from the Primary Care Evaluation of Mental Disorders (PRIME-MD) that measures symptoms of a major depression according to the *DSM-IV*. It consists of nine items that are answered on a 4-point Likert scale. Scores range from 0 to 27 points. Higher scores indicate more severe depressive symptoms. An additional item asks about the impact of symptoms on everyday life. The PHQ-9 is an efficient instrument to assess depression and shows good psychometric properties; its internal consistency is $\alpha = 0.86–0.89$ and test-retest reliability is $r = 0.84$ (Kroenke et al., 2001).

2.4.2. Rosenberg Self-Esteem Scale (RSE)

The Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965) is a 10-item self-report inventory assessing both positive and negative feelings about the self as levels of self-esteem. The items are rated on a 4-point Likert scale ranging from strongly agree to strongly disagree. The total score ranges from 10 to 40 points, in which higher scores represent higher self-esteem. The scale's internal consistency is $\alpha = 0.88$ (Roth et al., 2008). We used the German version by von Collani and Herzberg (2003).

2.4.3. WHO quality of life (WHOQOL-BREF)

The WHOQOL-BREF is an abbreviated 26-item version of the WHOQOL-100, which covers four domains of quality of life (QoL): physical, psychological, social, and environmental. We used the German version by Angermeyer et al. (2000). The total score ranges from 26 to 130 points. A high score indicates high quality of live. In a normal German population, the internal consistencies of the four subscales—physical, psychological, social, and environmental quality of life—ranged from $\alpha = 0.76$ to 0.88, and the discriminant validity was reported as good (Skevington et al., 2004).

2.4.4. University of Rhode Island Change Assessment (URICA)

Willingness to change was measured with an item subset of the University of Rhode Island Change Assessment (URICA) in the baseline survey, in a slightly adapted version. The URICA consists of the subscales precontemplation, contemplation, action, and maintenance. We used the German version by Hasler et al. (2003). We included this questionnaire because we expected willingness to change to be a moderator of the effectiveness of our self-help application. The subscales show internal consistencies ranging from $\alpha = 0.72$ to 0.86 (Hasler et al., 2003).

2.4.5. Client Satisfaction Questionnaire (CSQ-8)

The Client Satisfaction Questionnaire (CSQ-8; Attkisson and Zwick, 1982) was administered in the German version (Fragebogen zur Messung der Patientenzufriedenheit, ZUF-8; Schmidt et al., 1989) to participants who reported that they had used the application. We used an adapted version of the ZUF-8 to assess the participants' satisfaction with the application, such as its perceived quality, feasibility, effectiveness, and applicability to their problem and their intention to use the application in the future. The original version of the ZUF-8 assesses satisfaction after an inpatient treatment (Schmidt et al., 1989) and was adjusted for the application Be Good to Yourself (see Table 4).

2.5. Strategy of data analysis

Intention-to-treat (ITT), completer, and per protocol (PP) analyses

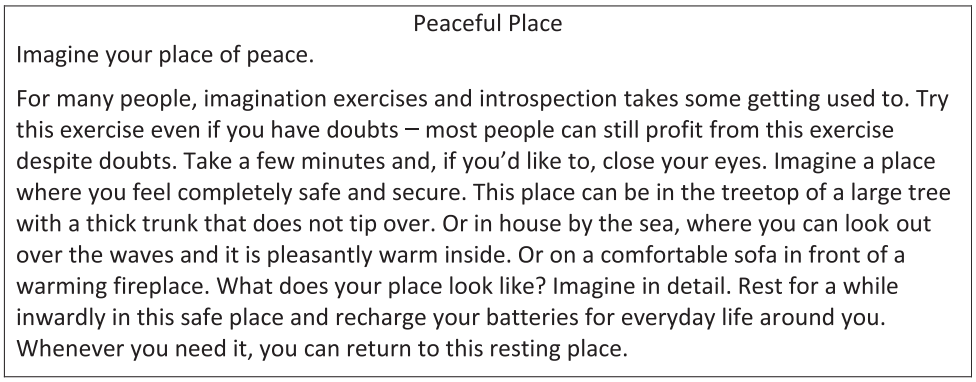

Fig. 3. Example of a task in the category mindfulness-based exercises.

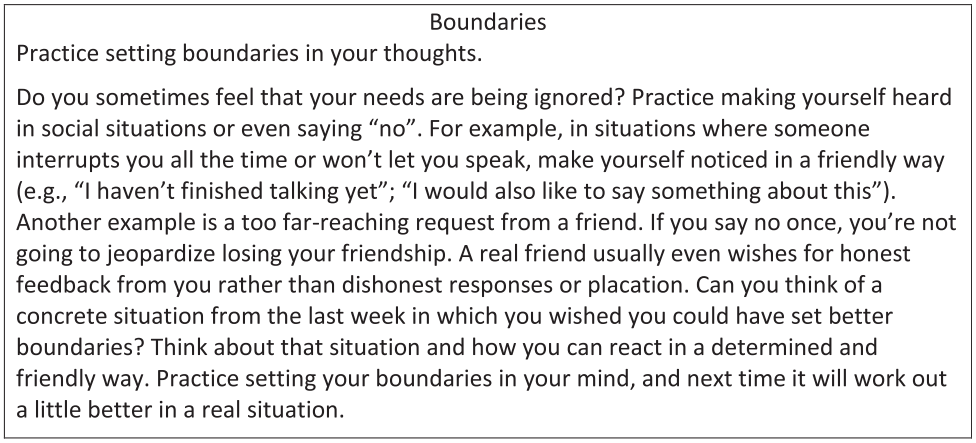

Fig. 4. Example of a task in the category social-competence skills.

were conducted. Completer analyses were carried out for participants with complete baseline and post data. Per protocol analyses were conducted for participants who used the application as requested (i.e., several times a week or more) and provided complete baseline and post data. ITT analyses considered data from all subjects with available baseline data. We conducted multiple imputation to estimate pre-post change scores for noncompleters (i.e., no data available at reassessment despite several reminders). The method of imputation was “fully conditional specification”, an iterative Markov chain Monte Carlo (MCMC) method, set as default by SPSS version 24, which assumes the pattern of missing data to be arbitrary. As imputation predictors, we used group allocation, gender, age, education, and therapy status (e.g., medication yes vs. no). We imputed 20 datasets (10 iterations) and report pooled p values for the effect of group allocation on the respective outcome. Main results were computed using ANCOVAs. For ITT and completer ANCOVAs, we report the group difference in pre-post change scores, correcting for baseline scores of the respective outcome. For per protocol ANCOVAs, gender is included as an additional covariate because subsamples differed significantly regarding this variable ($\chi^2 = 4.52$, $p = 0.03$). However, including two covariates considerably increases complexity to the model which might diminish results, considering the small sample size. Therefore, we additionally calculated a second, simpler model in which we corrected for gender only. We report effect sizes as partial eta squared ($\eta^2_{partial} \approx 0.01$ small effect, $\eta^2_{partial} \approx 0.06$ medium effect, $\eta^2_{partial} \approx 0.14$ large effect). Group differences at baseline were calculated using t tests and χ^2 tests. Additionally, moderator analyses were conducted with the SPSS plugin PROCESS (Hayes, 2013). We hypothesised that willingness to change at baseline would moderate the effect of group allocation on primary and secondary outcomes. All moderator models included group allocation as the independent variable, post scores of the outcome as the dependent variable, baseline scores of the outcome as a covariate, and willingness to change as a covariate, as well as the interaction between group allocation and willingness to change to test the moderation. A significant interaction indicates that willingness to change functions as a moderator, meaning that it influences the intervention's effectiveness compared to the wait-list control group. We used complete case data for all moderator analyses.

3. Results

3.1. Baseline differences

Baseline demographic and psychopathological characteristics of the treatment and the wait-list group are presented in Table 1. Randomization was successful: no significant differences between groups emerged for any of the demographic characteristics nor for the three psychopathological variables (depressive symptoms, self-esteem, and quality of life). Approximately three out of four participants were female. Symptom severity was moderate at baseline (see Table 2): 14.8% of the individuals ($n = 13$) met the criteria for severe depressive symptoms (PHQ-9 score 20–27), 18.2% ($n = 16$) had moderately severe depressive symptoms (PHQ-9 score 15–19), 28.4% ($n = 25$) had moderate depressive symptoms (PHQ-9 score 10–14); 28.4% ($n = 25$) had mild depressive symptoms (PHQ-9 score 5–9), and 10.2% ($n = 9$) met the criteria for no present depressive symptoms (PHQ-9 score 0–4). One of the inclusion criteria was the subjective wish to receive a treatment. Therefore, participants without current depressive symptoms but with, for example, a (subjective) risk of relapse, were included in the sample as well. Approximately half of the participants received psychotherapy and/or pharmacological treatment (see Table 1), which again was not different across groups at baseline, nor did the therapy status change differently between groups (change of psychotherapy: $t = 0.927$; $p = 0.356$; change of medication: $t = 0.899$; $p = 0.371$).

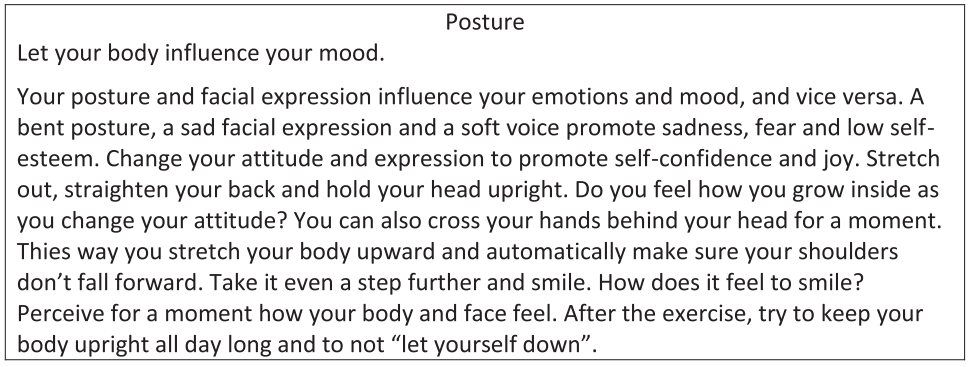

Fig. 5. Example of a task in the category activating exercises.

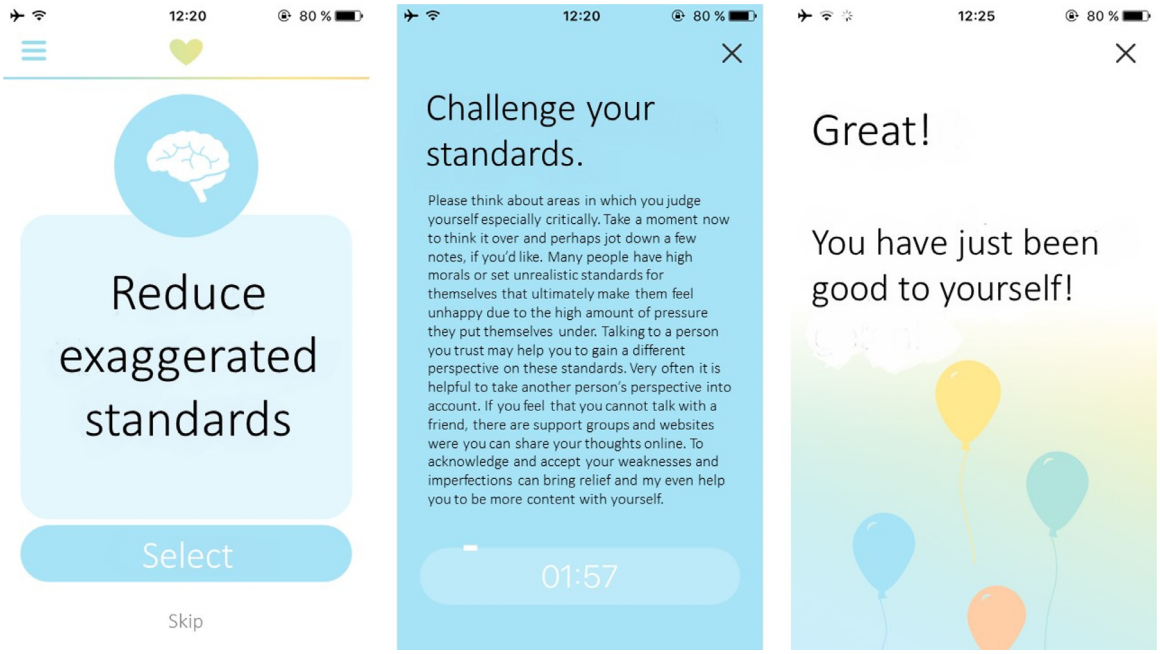

Fig. 6. Screenshots of the self-help application (f.l.t.r.): title screen; detailed description with a "finish" button that is only selectable after the countdown has finished; reinforcement screen.

3.2. Completion

The completion rate was 84%, and it did not differ between the treatment group and the wait-list group (see Table 1). Noncompleters were not significantly different from completers on baseline demographic and psychopathological variables except for gender, $\chi^2(1, 88) = 12.43$, $p < 0.001$; only 58% (11 out of 19) of the males completed both assessments while 91% (63 out of 69) of the females completed both assessments. The test-retest reliability of the outcome variables was satisfactory to good (PHQ-9: $r = 0.70$, $p < 0.001$; RSE: $r = 0.79$, $p < 0.001$; WHOQOL-BREF: $r = 0.88$, $p < 0.001$). Despite the high completion rate, only 39% of the participants in the intervention group actually used the application frequently (i.e., several times a week). Willingness to change was not correlated with self-reported frequency of usage ($r = 0.05$, $p = 0.805$).

3.3. Intention-to-treat and completer analyses

We conducted intention-to-treat (ITT) ANCOVAS for primary as well as secondary outcomes. All models included pre-post change scores as outcomes, with baseline scores of the respective outcome variable as covariates. The combined model-estimated marginal means of the pre-post change scores and corresponding standard errors (in brackets) are as follows. PHQ-9: change score = 1.70 (0.86) for the wait-list group vs. change score = 1.78 (0.87) for the intervention group; RSE: change score = −2.97 (1.12) for the wait-list group vs. change score = −1.20 (1.11) for the intervention group; WHOQOL-BREF: change score = −2.30 (1.43) for the wait-list group vs. change score = −1.57 (1.62) for the intervention group. Positive change scores indicate a decrease and negative change scores indicate an increase from pre to post assessment. The results, summarized in Table 2, show no difference between groups at all. Likewise, the completer analyses did not show any significant difference between groups (see Table 2). The combined model-estimated marginal means of the pre-post change scores and corresponding standard errors (in brackets) are as follows. PHQ-9: change score = 1.52 (0.69) for the wait-list group vs. change score = 1.94 (0.72) for the intervention group; RSE: change score = −2.89 (0.83) for the wait-list group vs. change score = −1.43 (0.88) for the intervention group; WHOQOL-BREF: change score = −2.04 (1.28) for the wait-list group vs. change score = −2.42 (1.37) for the intervention group. Subsidiary paired *t* test analyses showed that both the wait-list and intervention group improved over time on depressive symptoms, whereas only the wait-list group improved on the self-esteem scale over time (see Table 2).

3.4. Per protocol analyses

Participants of the wait-list group ($n = 39$) were compared to the per protocol users (i.e., those who used the application several times a week; $n = 19$) of the intervention group in a complete case analysis. For

Table 1
Demographic and medical background information, showing percentages, means, and standard deviations.

Background	Wait-list ($n = 44$)	Intervention group ($n = 44$)	Statistics (df)
Age in years	44.57 (10.69)	41.20 (11.86)	$t = 1.397$, $p = 0.166$ (86)
Gender (% in female)	75.00	81.82	$\chi^2 = 0.604$, $p = 0.437$ (1)
Education (13th grade) in %	56.81	72.73	$\chi^2 = 1.660$, $p = 0.646$ (3)
Psychotropic medication in %	50.00	45.45	$\chi^2 = 0.182$, $p = 0.669$ (1)
Wish to begin face-to-face therapy in %	15.91	13.64	$\chi^2 = 1.359$, $p = 0.224$ (1)
Completer in %	88.63	79.55	$\chi^2 = 0.090$, $p = 0.764$ (1)
Therapy status at baseline: in treatment by single therapist in %	38.64	59.09	$\chi^2 = 3.684$, $p = 0.055$ (1)
Therapy status at baseline: inpatient treatment in %	2.27	2.27	$\chi^2 = 0.000$, $p = 1$ (1)

Table 2
Completer and intention-to-treat analyses. Group differences across time on primary and secondary outcomes. Means and standard deviations (in brackets) are presented. Significant within-subject differences across time are displayed in square brackets.

Variables	Wait-list		Intervention		Completer between-group difference pre-post; ANCOVA	Intention-to-treat (multiple imputation) between-group difference pre-post; ANCOVA
	Pre (n = 44)	Post (n = 39)	Pre (n = 44)	Post (n = 35)		
PHQ-9	12.77 (6.40)	10.72 (6.05) [*]	11.61 (6.14)	10.23 (5.56) [*]	$F(1;71) = 0.173$, $p = 0.678$, $\eta_p^2 = 0.002$	$p = 0.952$
Rosenberg Scale	24.25 (7.57)	27.31 (8.32) [**]	24.98 (7.66)	25.66 (7.42)	$F(1;71) = 1.464$, $p = 0.230$, $\eta_p^2 = 0.020$	$p = 0.274$
WHOQOL-BREF	75.32 (14.55)	79.05 (15.76) [+]	79.40 (12.85)	80.76 (13.29)	$F(1;70) = 0.041$, $p = 0.840$, $\eta_p^2 < 0.001$	$p = 0.738$

Note. Significant difference from zero: + $p \leq 0.1$, *$p \leq 0.05$; ** $p \leq 0.01$, *** $p \leq 0.001$. Completer and Intention-to-treat analyses include baseline scores of the respective outcome as covariates. Intention-to-treat and completer pre-post change scores are presented in Section 3.3.

models with baseline scores and gender as covariates, the combined model-estimated marginal means of the pre-post change scores and corresponding standard errors (in brackets) are as follows: PHQ-9: change score = 1.68 (0.69) for the waitlist group vs. change score = 3.77 (1.01) for the intervention group; RSE: change score = −2.91 (0.90) for the waitlist group vs. change score = −2.92 (1.31) for the intervention group; WHOQOL-BREF: change score = −2.20 (1.25) for the waitlist group vs. change score = −4.76 (1.81) for the intervention group. For PHQ-9 change-scores, the groups differ at trend level with a small effect size ($p = 0.1$, $\eta_p^2 = 0.049$) in favor of the intervention group.

For models with gender as the only covariate, combined model-estimated marginal means of the pre-post change scores and corresponding standard errors (in brackets) are as follows: PHQ-9: change score = 1.47 (0.74) for the waitlist group vs. change score = 4.20 (1.08) for the intervention group; RSE: change score = −2.90 (0.92) for the waitlist group vs. change score = −2.94 (1.34) for the intervention group; WHOQOL-BREF: change score = −2.21 (1.24) for the waitlist group vs. change score = −4.73 (1.80) for the intervention group. The results of the per protocol analyses are summarized in Table 3. When only including gender as covariate, the group difference reaches statistical significance with a medium effect in favor of the intervention group ($p = 0.045$, $\eta_p^2 = 0.071$). There were no differences between groups regarding secondary outcomes.

3.5. Subjective benefit

Table 4 provides data on the subjective evaluation and satisfaction with the self-help application (this includes only the participants in the intervention group who used the application). The majority of participants evaluated the self-help application as positive and intended to use the application in the future.

3.6. Moderator analyses

Willingness to change did not moderate the effect of group allocation on post-treatment depression scores, self-esteem, quality of life, or psychological quality of life (all p values ≥ 0.35). Therefore, participants' willingness to change did not influence whether the intervention was effective compared to a wait-list control group.

4. Discussion

The study investigated the feasibility and efficacy of a novel smartphone application for depressive symptoms, Be Good to Yourself, which is currently available in German for the iPhone operating system. It was expected that depressive symptoms would decrease with the use of the self-help application and that self-esteem and quality of life would increase.

4.1. Primary outcome

Depressive symptoms decreased in both groups during the intervention time of four weeks, therefore no significant difference between the treatment group and the control group emerged for the primary outcome. The improvement in the control group might be attributable to the passing of time, self-efficacy, patients' use of treatments other than Be Good to Yourself (see Table 1), or spontaneous remission (Whiteford et al., 2013).

The high rate of participants who did not use the application on a regular basis was striking. Determining the reasons for those high rates should be a high priority in future research projects (e.g., is e-mental health less binding than face-to-face therapy or does e-mental health overburden users?). This high rate is congruent with other studies (Christensen et al., 2009; Richards and Richardson, 2012; Titov, 2011) and seems to be one of the biggest barriers to online interventions in

Table 3
Per protocol analyses. Group differences across time on primary and secondary outcomes. Means and standard deviations (in brackets) are reported for completers. Significant within-subject differences across time are displayed in square brackets.

Variables	Wait-list		Intervention		Per protocol between-group difference pre-post; ANCOVA
	Pre (n = 39)	Post (n = 39)	Pre (n = 19)	Post (n = 19)	
PHQ-9	12.26 (6.04)	10.72 (6.05)	14.11 (5.37)	10.05 (5.24) [***]	[a] $F(1;55) = 2.793$, $p = 0.100$, $\eta_p^2 = 0.049$ [b] $F(1;55) = 4.193$, $p = 0.045$, $\eta_p^2 = 0.071$
Rosenberg Scale	24.44 (7.70)	27.31 (8.32)	23.47 (6.05)	26.47 (6.92) [+]	[a] $F(1;55) < 0.001$, $p = 0.998$, $\eta_p^2 < 0.001$ [b] $F(1;55) < 0.001$, $p = 0.985$, $\eta_p^2 < 0.001$
WHOQOL-BREF	76.97 (13.33)	79.05 (15.76)	78.53 (9.99)	83.53 (10.11) [*]	[a] $F(1;55) = 1.306$, $p = 0.258$, $\eta_p^2 = 0.024$ [b] $F(1;55) = 1.287$, $p = 0.262$, $\eta_p^2 = 0.023$

Note. Significant difference from zero: + $p \leq 0.1$, *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$.
[a] Per protocol analyses with gender and baseline scores as covariates.
[b] Per protocol analyses with gender as covariate.

Table 4
Client Satisfaction Questionnaire CSQ-8 (German Version ZUF-8) (n = 26).

Items	$\bar{x}$	% positive	s; range
1. How would you rate the quality of the application you received? (“Excellent”, “Good” vs. “Fair”, “Poor”)	1.85	88.5%	0.61; 1–3
2. Did you get the kind of help you wanted? (“Yes, definitely”, “Yes, generally” vs. “No, not really”, “No, definitely not”)	2.27	65.4%	0.83; 1–4
3. To what extent has our Application met your needs? (“Almost all of my needs have been met”, “Most of my needs have been met” vs. “Only a few of my needs have been met”, “None of my needs have been met”)	2.35	57.7%	0.85; 1–4
4. If a friend needed similar help, would you recommend our program to him/her? (“Yes, definitely”, “Yes, I think so” vs. “No, I don't think so”, “No, definitely not”)	2.08	73.1%	0.89; 1–4
5. How satisfied are you with the amount of help you received from the application (“Very satisfied”, “Mostly satisfied” vs. “Indifferent or mildly dissatisfied”, “Quite dissatisfied”)	2.35	61.5%	0.89; 1–4
6. Has the application helped you to deal more effectively with your problems? (“Yes, it helped a great deal”, “Yes, it helped somewhat”, vs. “No, it really didn't help”, “No, it seemed to make things worse”)	2.23	65.4%	0.65; 1–3
7. In an overall, general sense, how satisfied are you with the application you received? (“Very satisfied”, “Mostly satisfied” vs. “Indifferent or mildly dissatisfied”, “Quite dissatisfied”)	2.31	61.5%	0.93; 1–4
8. If you were to seek help again, would you use our application again? (“Yes, definitely”, “Yes, I think so” vs. “No, I do not think so”, “No, definitely not”)	2.12	69.2%	0.99; 1–4

Note. The rate for % positive counted the responses for the two positive options stated in brackets in the first column.

general, that have an average non-completion rate of 31% (Melville et al., 2010). The high attrition rate can possibly be overcome by combining the smartphone application with classical psychotherapy, as studies have shown that adding online sessions to face-to-face interventions (i.e., blended therapy) can increase treatment adherence (van der Vaart et al., 2014).

In per protocol analyses, the treatment group improved in depressive symptoms at trend level with a small effect size compared with the wait-list group when correcting for baseline scores and gender. When correcting for gender only, this difference reached statistical significance at a medium effect size. This last finding is congruent with results of a meta-analysis on the effectiveness of unguided online self-help treatments (Karyotaki et al., 2017), although in our study the result is restricted to the per protocol group only. The meta-analysis included studies that investigated more time-consuming and extensive web-based interventions. Therefore, the only partially significant results found in our study could be due to the brevity of the smartphone application (the short exercises). The sample size of the treatment group in the per protocol analyses was small (n = 19) which limits the generalizability. Additionally, studies show that more severely affected participants seem to benefit more from low intensity interventions for depression (Bower et al., 2013), and as the treatment group in the per protocol analyses was more severely depressed at baseline, our results only reached statistical significance when no correction for baseline depression was applied. Further research is needed to substantiate our per protocol results. Because of the high number of participants who did not use the application frequently, the issue of how to better motivate participants deserves further research. A posthoc analysis revealed that willingness to change was not correlated with self-reported frequency of usage, but there may be other factors that influence how frequently an application is used. For this study, the application presented self-help tasks in a standardized order. It should be evaluated whether the frequency of usage can be increased via the opportunity to save preferred tasks on a list and whether the implementation of gamification elements would motivate participants to use the intervention more often. Gamification is the usage of game design components and game principles in a nongame context, for example, by using a visual reward system (for detailed information see Deterding et al., 2011).

4.2. Secondary outcomes

Both of the secondary outcomes, that is, self-esteem (RSE) and quality of life (WHOQOL-BREF), showed no difference between groups, neither in the ITT, completer, nor per protocol analyses. Although participants improved overall in the two secondary measures between pre and post, the results did not indicate improved self-esteem or quality of life due to the self-help application.

4.3. Feasibility

The results of the Client Satisfaction Questionnaire (ZUF-8) suggest that users evaluated the application positively. Especially striking is the overall positive evaluation of the quality of the application, the high percentage of participants who would recommend the application, and the high percentage of participants who would use it in the future. However, the results indicate that the application did not meet the needs of almost half of the participants, which leads to the conclusion that it should be further modified. To generate ideas on how to improve the application we inspected the feedback provided by our participants. Several participants mentioned that the countdown, which we implemented to ensure that participants took enough time for a task, was too slow or perceived as unnecessary. This is an important point of critique as it illustrates how a feature can lead to frustration although it was intentionally implemented to improve users’ experiences. Other participants criticized the quality of audio files. Yet another critical feedback concerned the reward message that participants received after completing a task (reinforcement screen, see Fig. 6). As the message was always identical, one participant reported that seeing the reinforcement screen was not experienced as rewarding anymore after reading it several times.

Since participants answered all questions on satisfaction positively, with more than 50% approval (we set the threshold criterion for feasibility a priori at 50%), our application seems to be a feasible intervention to address mild to moderate depressive symptoms. The self-help application should be optimized, however, and it should be adapted for other operating systems so it can reach more people.

4.4. Strengths and limitations

The research project was conducted with a limited budget and therefore limited technical expertise. Because of these reasons, the self-help application was only developed for the operating system iOS, which runs on iPhones. The sample for this study therefore includes only iPhone users. In future research projects, self-help smartphone applications should be evaluated for a wider variety of operating systems. Also, groups differed at trend level regarding participants’ therapy status (see Table 1). In the intervention group, more participants received therapy from an individual therapist which might have led to an overestimation of treatment effects.

A diagnosis of depression was not an inclusion criterion, which means that the sample was very heterogeneous regarding depression severity. On the one hand, this is a strength of the study since we reached individuals who subjectively felt the need for psychological treatment, irrespective of diagnosis, which may have led to a representative sample. On the other hand, research shows that

participants with more severe symptoms consistently benefit more from psychological interventions. Thus, the broad inclusion criteria may have increased the likelihood of a type I error, resulting in an underestimation of the intervention's true potential. Another strength of this project is the high completion rate even though participants did not use the application frequently.

5. Conclusion

Smartphone-based self-help applications could represent a promising complementary tool to address depressive symptoms in individuals who show high adherence in terms of frequent usage. Also, patients who don't have access to a computer might benefit from such applications. Our intervention was only effective at trend level in frequent users (per protocol). We are currently investigating how the efficacy of the application can be augmented, such as by providing the possibility of selecting favorite exercises that would be presented more often or by implementing gamification elements that might motivate users. Moreover, it has not yet been established whether the magnitude of the effect could be enhanced through blended treatment.

Acknowledgments

The research project was not externally funded and was planned and executed independent of any conflicting interests.

The authors Lilian Pult and Steffen Moritz developed the application, which was evaluated in the present study. With publication of this paper, no income has been generated with the application.

The authors report no conflicts of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2018.08.113.

References

Andersson, E., Steneby, S., Karlsson, K., Ljótsson, B., Hedman, E., Enander, J., et al., 2014. Long-term efficacy of Internet-based cognitive behavior therapy for obsessive-compulsive disorder with or without booster: a randomized controlled trial. Psychol. Med. 44 (13), 1–11.

Angermeyer, M.C., Kilian, R., Matschinger, H., 2000. WHOQOL-100 Und QHOQOL-BREF: Handbuch für die Deutschsprachige Version Der WHO Instrumente zur Erfassung von Lebensqualität. Hogrefe.

Attkisson, C.C., Zwick, R., 1982. The Client Satisfaction Questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval. Prog. Plan. 5 (3), 233–237.

Bert, F., Giacometti, M., Gualano, M.R., Siliquini, R., 2014. Smartphones and health promotion: a review of the evidence. J. Med. Syst. 38 (1). https://doi.org/10.1007/s10916-10013-19995-10917.

Boswell, J.F., Sauer-Zavala, S.E., Gallagher, M.W., Delgado, N.K., Barlow, D.H., 2012. Readiness to change as a moderator of outcome in transdiagnostic treatment. Psychother. Res. 22 (5), 570–578.

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D.A., Gilbody, S., et al., 2013. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ 346, f540 February.

Christensen, H., Griffiths, K.M., Farrer, L., 2009. Adherence in internet interventions for anxiety and depression: systematic review. J. Med. Internet Res. 11 (2), e13.

Churchill, R., Moore, T.H., Furukawa, T.A., Caldwell, D.M., Davies, P., Jones, H., et al., 2013. 'Third wave' cognitive and behavioural therapies versus treatment as usual for depression. Cochrane Database Syst. Rev.(10), CD008705.

Cuijpers, P., Donker, T., Johansson, R., Mohr, D.C., van Straten, A., Andersson, G., 2011. Self-guided psychological treatment for depressive symptoms: a meta-analysis. PLoS One 6 (6), e21274.

Deterding, S., Dixon, D., Khaled, R., Nacke, L., 2011. From game design elements to gamefulness: defining "gamification". Proceedings of the 15th international academic mindtrek conference: envisioning future media environments ACM, New York, NY, USA, pp. 9–15. https://doi.org/10.1145/2181037.2181040.

Driessen, E., Smits, N., Dekker, J.J., Peen, J., Don, F.J., Kool, S., et al., 2016. Differential efficacy of cognitive behavioral therapy and psychodynamic therapy for major depression: a study of prescriptive factors. Psychol. Med. 46 (4), 731–744.

Ebert, D., Donkin, L., Andersson, G., Andrews, G., Berger, T., Carlbring, P., et al., 2016. Does internet-based guided-self-help for depression cause harm? An individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials. Psychol. Med. 46 (13), 2679–2693.

Eichenberg, C., Wolters, C., Brahler, E., 2013. The internet as a mental health advisor in Germany–results of a national survey. PLoS One 8 (11), e79206.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., et al., 2017. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 16 (3), 287–298.

Griffiths, K.M., Farrer, L., Christensen, H., 2010. The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials. Med. J. Aust. 192 (11), 4–11.

Hasler, G., Klaghofer, R., Buddeberg, C., 2003. Der Fragebogen zur Erfassung der Veränderungsbereitschaft (FEVER). PPmP 53 (9/10), 406–411.

Hayes, A.F., 2013. Introduction to Mediation, Moderation, and Conditional Process Analysis: a Regression-Based Approach. Guilford Press, New York.

Heider, J., Köck, K., Sehlbrede, M., Schröder, A., 2018. Readiness to change as a moderator of therapy outcome in patients with somatoform disorders. Psychother. Res. 28 (5), 722–733.

Jelinek, L., Hauschildt, M., Moritz, S., 2015. Metakognitives Training bei Depression (D-MKT): Mit E-Book Inside Und Trainingsmaterial. Beltz, J, Weinheim.

Jelinek, L., Hauschildt, M., Wittekind, C.E., Schneider, B.C., Kriston, L., Moritz, S., 2016. Efficacy of metacognitive training for depression: a randomized controlled trial. Psychother. Psychosom. 85 (4), 231–234.

Johansson, R., Andersson, G., 2012. Internet-based psychological treatments for depression. Expert Rev. Neurother. 12 (7), 861–870.

Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., et al., 2017. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 74 (4), 351–359.

Kazdin, A.E., 2017. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. Behav. Res. Ther. 88, 7–18.

Kazdin, A.E., Blase, S.L., 2011. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect. Psychol. Sci. 6 (1), 21–37.

Khoury, B., Lecomte, T., Fortin, G., Masse, M., Therien, P., Bouchard, V., et al., 2013. Mindfulness-based therapy: a comprehensive meta-analysis. Clin. Psychol. Rev. 33 (6), 763–771.

Kohn, R., Saxena, S., Levav, I., Saraceno, B., 2004. The treatment gap in mental health care. Bull. World Health Organ. 82 (11), 858–866.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The phq-9. J. Gen. Intern. Med. 16 (9), 606–613.

Löwe, B., Spitzer, R., Zipfel, S., Herzog, W., 2002. Gesundheitsfragebogen für Patienten (PHQ-D). Manual und Testunterlagen. 2. Auflage. PRIME MD Patient Health Questionnaire (PHQ)–German version. Manual and Materials, 2nd ed. Pfizer, Karlsruhe.

Ly, K.H., Topooco, N., Cederlund, H., Wallin, A., Bergström, J., Molander, O., et al., 2015. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 10 (5), 1–16.

Melville, K.M., Casey, L.M., Kavanagh, D.J., 2010. Dropout from internet-based treatment for psychological disorders. Br. J. Clin. Psychol. 49 (4), 455–471.

Mohr, D.C., Ho, J., Duffecy, J., Baron, K.G., Lehman, K.A., Jin, L., et al., 2010. Perceived barriers to psychological treatments and their relationship to depression. J. Clin. Psychol. 66 (4), 394–409.

Normann, N., Van Emmerik, A.A.P., Morina, N., 2014. The efficacy of metacognitive therapy for anxiety and depression: a meta-analytic review. Depress. Anxiety 31 (5), 402–411.

Palfai, T.P., Tahaney, K., Winter, M., Saitz, R., 2016. Readiness-to-change as a moderator of a web-based brief intervention for marijuana among students identified by health center screening. Drug Alcohol Depend 161, 368–371.

Poushter, J., 2016. Smartphone ownership and internet usage continues to climb in emerging economies. http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ (accessed 27 April 2018).

Prochaska, J.O., DiClemente, C.C., 1982. Transtheoretical therapy: toward a more integrative model of change. Psychother. Theor. Res. Pract. 19 (3), 276–288.

Richards, D., Richardson, T., 2012. Computer-based psychological interventions for depression treatment: a systematic review and meta-analysis. Clin. Psychol. Rev. 32, 329–342.

Rosenberg, M., 1965. Society and the Adolescent Self-Image, Revised edition. Wesleyan University Press, Middletown, CT.

Roth, M., Decker, O., Herzberg, P.Y., Brähler, E., 2008. Dimensionality and norms of the Rosenberg self-esteem scale in a German general population sample. Eur. J. Psychol. Assess. 24 (3), 190–197.

Schmidt, J., Lamprecht, F., Wittmann, W., 1989. Zufriedenheit mit der stationären Versorgung. Entwicklung eines Fragebogens und erste Validitätsuntersuchungen. PPmP 39 (7), 248–255.

Schröder, J., Berger, T., Westermann, S., Klein, J.P., Moritz, S., 2016. Internet interventions for depression: new developments. Dialogues Clin. Neurosci. 18 (2), 203–212.

Skevington, S.M., Lotfy, M., O'Connell, K.A., 2004. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual. Life Res. 13 (2), 299–310.

Strauss, C., Cavanagh, K., Oliver, A., Pettman, D., 2014. Mindfulness-based interventions for people diagnosed with a current episode of an anxiety or depressive disorder: a meta-analysis of randomised controlled trials. PLoS One 9 (4), e96110.

Titov, N., 2011. Internet-delivered psychotherapy for depression in adults. Curr. Opin. Psychiatry 24 (1), 18–23.

Twomey, C., Reilly, G.O., Meyer, B., 2017. Effectiveness of an individually-tailored computerised CBT programme (Deprexis) for depression: a meta-analysis. Psychiatry

Res 256, 371–377. https://doi.org/10.1016/j.psychres.2017.06.081.
van der Vaart, R., Witting, M., Riper, H., Kooistra, L., Bohlmeijer, E.T., van Gemert-Pijnen, L.J., 2014. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 14 (1). doi:10.1186/s12888-014-0355-z.
Vandeleur, C.L., Fassassi, S., Castelao, E., Glaus, J., Strippoli, M.F., Lasserre, A.M., et al., 2017. Prevalence and correlates of DSM-5 major depressive and related disorders in the community. Psychiatry Res 250, 50–58. https://doi.org/10.1016/j.psychres.2017.01.060.
von Collani, G., Herzberg, P.Y., 2003. Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg. Zeitschrift für differentielle und diagnostische Psychologie 24 (1), 3–7.
Warmerdam, L., Smit, F., Van Straten, A., Riper, H., Cuijpers, P., 2010. Cost-utility and cost-effectiveness of Internet-based treatment for adults with depressive symptoms: randomized Trial. J. Med. Internet Res. 12 (5), 1–11.
Wells, A., Papageorgiou, C., 2004. Metacognitive therapy for depressive rumination. In: Papageorgiou, C., Wells, A. (Eds.), Depressive Rumination: Nature, Theory and Treatment. Wiley, West Sussex, pp. 259–273.
Whiteford, H., Harris, M., McKeon, G., Baxter, A., Pennell, C., Barendregt, J., et al., 2013. Estimating remission from untreated major depression: a systematic review and meta-analysis. Psychol. Med. 43 (8), 1569–1585.
Young, K.S., 2005. An empirical examination of client attitudes towards online counseling. Cyberpsychol. Behav. 8 (2), 172–177.