



Published in final edited form as:

Behav Ther. 2014 May ; 45(3): 358–375. doi:10.1016/j.beth.2014.02.001.

Feasibility Study of an Interactive Multimedia Electronic Problem Solving Treatment Program for Depression: A Preliminary Uncontrolled Trial

Margit I. Berman,

Geisel School of Medicine at Dartmouth

Jay C. Buckey Jr.,

Geisel School of Medicine at Dartmouth

Jay G. Hull,

Geisel School of Medicine at Dartmouth

Eftihia Linardatos,

Geisel School of Medicine at Dartmouth

Sueyoung L. Song,

Geisel School of Medicine at Dartmouth

Robert K. McLellan, and

Geisel School of Medicine at Dartmouth

Mark T. Hegel

Center for Technology and Behavioral Health Geisel School of Medicine at Dartmouth

Abstract

Computer-based depression interventions lacking live therapist support have difficulty engaging users. This study evaluated the usability, acceptability, credibility, therapeutic alliance and efficacy of a stand-alone multimedia, interactive, computer-based Problem Solving Treatment program (*ePST*TM) for depression. The program simulated live treatment from an expert PST therapist, and delivered 6 *ePST*TM sessions over 9 weeks. Twenty-nine participants with moderate-severe symptoms received the intervention; 23 completed a minimally adequate dose of *ePST*TM (at least 4 sessions). Program usability, acceptability, credibility, and therapeutic alliance were assessed at treatment midpoint and endpoint. Depressive symptoms and health-related functioning were assessed at baseline, treatment midpoint (4 weeks), and study endpoint (10 weeks). Depression outcomes and therapeutic alliance ratings were also compared to previously published research on live PST and computer-based depression therapy. Participants rated the program as highly usable, acceptable, and credible, and reported a therapeutic alliance with the program comparable to that observed in live therapy. Depressive symptoms improved significantly over

time. These findings also provide preliminary evidence that ePST™ may be effective as a depression treatment. Larger clinical trials with diverse samples are indicated.

Keywords

Electronic Problem Solving Treatment; depression treatment; computer-based therapy; therapeutic alliance

The lifetime prevalence for major depressive disorder in the United States is approximately 15% (Kessler et al., 2003). Many depressed individuals are undertreated or receive no treatment. In the National Comorbidity Study Replication, only 56% of participants with a mood disorder had sought treatment, and only 38% of those received minimally adequate treatment (Wang et al., 2005). Inaccessible evidence-based treatments (especially psychotherapy), inconvenient clinic hours, distant locations, high costs, concerns about stigma, and worries about privacy all prevent depressed individuals from accessing treatment (Hirschfeld, Keller, Panico & Arons, 1997).

To reduce these barriers, computer-based (CB) depression treatments have been developed, both as stand-alone interventions and integrated into psychotherapy (Cavanagh et al., 2006; Kaltenthaler et al., 2006; Proudfoot et al., 2003). Individuals with little time, or who value privacy and independence, prefer CB therapy to live therapy (Mohr et al., 2010). In addition, because CB therapies are inexpensive, using them could increase access to care (Andersson & Cuijpers, 2009).

CB therapy can be effective for treating depression, with small to moderate effect sizes (Andersson & Cuijpers, 2009). Dropout rates are high, however—57% in CB therapies for depression (Richards & Richardson, 2012) compared to an average dropout rate of 20% in face-to-face psychotherapy (Swift & Greenberg, 2012). Participants may drop out of CB therapies for many reasons, including lack of acceptability or poor efficacy. One explanation for the high dropout rate may be the program's inability to engage users and develop a therapeutic alliance with them, which may also decrease acceptability or efficacy. Many CB therapies are text-dominated, noninteractive programs that fail to provide tailored feedback to the user. These programs may not develop the affective bond between therapist and client that underpins the therapeutic alliance (Martin, Garske, & Davis, 2000; Ormrod, Kennedy, Scott, & Cavanagh, 2010; Zuroff et al., 2000). Consistent with this, CB treatments for depression lacking human support have higher dropout rates than those that have such support (Richards & Richardson, 2012).

Although CB treatments that include some interaction with a mental health professional develop alliance comparable to that observed in face-to-face therapy (Reynolds, Stiles & Grohol, 2006; Socala et al., 2012), the therapeutic alliance is not as predictive of outcome in CB therapies compared to live psychotherapy (Andersson et al., 2012; Knaevelsrud & Maercker, 2006). Also, CB treatments provide poorer alliance than other treatment delivery methods (King, Bambling, Reid & Thomas, 2006; Leibert & Archer, 2006). At the same time, CB treatment models vary in ways relevant to forming a therapeutic alliance, such as the program's ability to interact with the user based on their inputs, involvement of mental

health professionals, ability to interact with peers who share the users' concerns, and use of audiovisual media.

One factor known to influence CB treatment outcome is professional support. CB researchers have identified several gradations of professional support, including (a) self-administered (SA) treatments (no therapist or nonprofessional contact beyond research assessment); (b) self-administered, structured (SAS) treatments (the participant must make an appointment to use the program in a lab or clinic setting, and may interact briefly with non-professionals); (c) predominantly self-help (PSH) interventions (therapists contact participants briefly to train them in using the CB treatment or provide a treatment rationale); (d) minimal-contact (MC) therapy (therapists actively assist the client in applying therapy techniques), and (e) therapist-administered (TA) treatment (face-to-face psychotherapy in which a CB treatment tool is used to enhance the therapy; Newman, Erickson, Przeworski & Dzus, 2003; Newman, Szkodny, Llera, & Przeworski, 2011).

CB depression treatment programs offering live professional support are more effective than those that do not, suggesting that programs without such support do a relatively poor job of creating a therapeutic alliance with users (Andersson & Cuijpers, 2009; Richards & Richardson, 2012; Spek et al., 2007). But, integrating professional support into a CB treatment increases both the expense and the logistical challenges of providing the treatment. Therefore, an SA or SAS CB depression treatment that could generate an effective therapeutic alliance without requiring professional support might improve access to depression care significantly. Stand-alone CB depression treatments can enhance the therapeutic alliance with users in many ways, such as by creating user-accessible programs, generating belief in the program's helpfulness, providing empathy by offering insights into the depression experience, conveying warmth via audio or video recordings, or empowering users to access self-help resources within the program (Barazzone, Cavanagh, & Richards, 2012). Although some of the commonly used, empirically supported stand-alone CB depression treatments (e.g., Beating the Blues, MoodGYM, Living Life to the Full; Proudfoot et al., 2004; Pittaway et al., 2009; Christensen, Griffiths & Korten, 2002) incorporate features recommended for establishing a therapeutic alliance with users (e.g., providing positive expectancies and hope; Richardson, Richards, & Barkham, 2010), they are not as robust at providing features that develop the relationship over time (e.g., developing commitment and trust) or that maintain the relationship (e.g., responsiveness to difficulties experienced and flexibility in response to user inputs; Barazzone et al., 2012).

In the current study, we evaluated a multimedia-based, immersive, computer-based Problem Solving Treatment program (*ePST*TM) designed to create a stand-alone therapeutic alliance that responds flexibly to users' experience over time. The *ePST*TM program was built for the U.S. National Aeronautics and Space Administration (NASA) as part of a self-guided, confidential, psychosocial training program intended to help astronauts on long missions (Carter, Buckey, Greenhalgh, Holland, & Hegel, 2005). *ePST*TM uses problem-solving therapy (Nezu, D'Zurilla, Zwick, & Nezu, 2004), and was adapted from Problem-Solving Treatment for Primary Care (PST-PC; Arean, Hegel, Vannoy, Fan & Unutzer, 2008; Mynors-Wallis, Gath, Day, & Baker, 2000; Mynors-Wallis, Gath, Lloyd-Thomas & Tomlinson, 1995). Therapeutic alliance is enhanced by using a virtual therapist (MTH) who

appears onscreen throughout the user's engagement with *ePST*TM and who interacts flexibly via branching algorithms that provide personalized responses from the therapist to the user based on their input about their problem-solving experiences and depressive symptoms. The *ePST*TM program is particularly strong in providing flexible troubleshooting assistance when the user encounters difficulty with their problem-solving efforts. This feature is intended to promote the development and maintenance of the therapeutic alliance over time, a weakness noted for other CB depression programs (Barazzone et al., 2012).

Although several stand-alone CB depression treatments are narrated by a mental health professional (e.g., Pittaway et al., 2009) or feature videos of model patients to normalize the experience of depression (e.g., Proudfoot et al., 2004), we are not aware of any CB programs for depression that use a virtual therapist that interacts directly and flexibly with users. The *FitTrack* program, a stand-alone CB intervention to enhance exercise and behavioral health, included a cartoon avatar trainer, which was programmed to engage in interactive text-based chats with users and to use relational behaviors, including empathy, social dialogue, and nonverbal immediacy. Versions of the *FitTrack* program that included the empathic interactive avatar led to greater working alliance and increased engagement with the program than versions that included no avatar or a "nonrelational" avatar that did not emit relationship-enhancing behaviors (Bickmore, Gruber, & Picard, 2005), suggesting that a virtual helper can be created that enhances the therapeutic alliance of stand-alone CB interventions. If *ePST*TM is feasible, efficacious, and able to create a robust therapeutic alliance, it will represent a step forward in the development of stand-alone CB depression treatments.

A previous pilot trial of the *ePST*TM program (Cartreine, Locke, Buckey, & Hegel, 2012) suggested that the program was credible, acceptable, and usable in a small group of participants with minor depression.¹ In the present study, we evaluated the program in a larger sample of mildly to moderately depressed adults to assess the credibility, acceptability, usability, and therapeutic alliance of *ePST*TM for depression. We also collected preliminary data on clinical outcomes from *ePST*TM, including improvement in depressive symptoms. We hypothesized that (1) Participants would rate *ePST*TM highly credible, acceptable and usable; (2) The program would create a strong therapeutic alliance with participants; and (3) Depression would show clinically and statistically significant improvement over time. To put our results in context, we conducted exploratory analyses to compare our findings to those obtained in other trials of both computer-based and live psychotherapy. Specifically, we compared (a) the therapeutic alliance developed with *ePST*TM to that obtained in studies of both live cognitive-behavioral therapy for depression and another CB depression treatment program, and (b) depression outcomes from *ePST*TM with outcomes using therapist-directed PST in a previous trial conducted by our research group for minor depression (Oxman, Hegel, Hull, & Dietrich, 2008), as well as a previous trial evaluating live PST with a population similar to ours for depression severity (Mynors-Wallis et al., 1995).

¹Note that none of the participants or data from this earlier study overlap with the participants or data presented here.

Methods

The Institutional Review Board of Dartmouth College approved this study.

PARTICIPANTS

Participants were recruited via referral from the Employee Assistance Program (EAP) of a large academic medical center, the EAP of a liberal arts university in rural New England, a health coaching program at the medical center, posters, and newspaper advertisements. All participants were recruited between June 1, 2011, and March 15, 2012. The sample size was determined by the participants we were able to recruit during that time.

Participants were included if they were age 18 or older; had a Hamilton Depression Rating Scale (HAMD) score greater than 10 (indicating at least mild depression); had two or more symptoms of depression as defined by the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV; American Psychiatric Association, 1994), one of which was depressed mood or lack of pleasure; had persistent depressive symptoms for at least 4 weeks; were either willing to postpone the use of an anti-depressant medication or were on a stable antidepressant regimen (no medication changes in the past 8 weeks or during study participation); were not currently receiving and were willing to postpone receiving psychotherapy; and were familiar with personal computers. Participants were excluded if they had a history of schizophrenia, bipolar disorder, obsessive-compulsive disorder, anorexia nervosa, or bulimia nervosa; had current active suicidal ideation or a recent (within the past 10 years) history of a suicide attempt or self-injurious behavior; or met criteria for current substance abuse or dependence. We excluded participants with these characteristics to avoid including participants who could not safely forego access to other psychiatric treatment during study participation, and to improve the homogeneity of the sample.

The average age of the sample was 53 years ($SD = 12.6$). They were mostly female (79%) and non-Hispanic White (97%); one participant was Native American. The majority were married (55%), employed (66%), had attended college (83%), and had an annual income greater than \$40,000 (52%). Participants began the study significantly depressed: 80% met criteria for major depressive disorder at the initial assessment; 17% met criteria for dysthymic disorder (as assessed by Structured Clinical Interview for DSM-IV; First, Spitzer, Gibbon, & Williams, 2002). The median number of previous depressive episodes reported by participants was 10, and 62% were taking antidepressant medications at the initial assessment.

PROCEDURE

Following prescreening, eligible participants had an initial assessment, where informed consent was obtained and demographic, eligibility, and baseline measures were completed. Participants were scheduled to complete 6 *ePST*TM sessions over 9 weeks using a laptop computer and printer in the library of the medical center, individually, in a private study room. Because the program is highly interactive and varies in content depending on users' choices, sessions varied in length, but generally lasted 30 minutes to 1 hour. A nonclinician study coordinator was available if participants had questions while using the program. The

coordinator also reviewed participants' Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001) responses at each *ePST*TM session to check for suicidality. If participants endorsed suicidality, a study psychologist completed a safety assessment within 24 hours. During the trial 18 such assessments for suicidality were conducted. In no case did a participant report an active intent or plan to self-harm, and consequently all participants who completed the safety assessment continued in the study. Study outcome assessments were administered at Weeks 4 and 10. Participants received a \$20 cash incentive for completing each of the baseline, Week 4, and Week 10 assessments.

The *EPST*TM PROGRAM

At each *ePST*TM session, when users logged in, they were welcomed by a therapist (MTH), presented via audio and video. The user then completed the computer-administered PHQ-9, and the virtual therapist provided personalized feedback about symptoms and progress. The first *ePST*TM session provided education about depression and the six steps of problem-solving therapy:

1. Clarifying the problem.
2. Establishing an achievable goal.
3. Brainstorming solutions.
4. Evaluating the pros and cons of each solution and selecting one or more solutions.
5. Developing an action plan to implement the solution(s).
6. Evaluating the success of the implementation and troubleshooting as needed.

During all sessions, the virtual therapist guided users through all steps of problem-solving therapy for at least one problem, plus scheduling enjoyable activities. The therapist also provided personalized feedback about users' progress solving their problems and improving depression, and helped users troubleshoot difficulties. At the end of each *ePST*TM session, users received a printout summarizing the work on each problem, the action plan to be completed before the next session, and a schedule of their self-generated enjoyable activities for the coming week.

MEASURES

All clinician-rated measures were administered by either a licensed psychologist (MIB, SS) or a postdoctoral fellow with a Ph.D. in clinical psychology (EL). All raters had previous experience administering and scoring structured clinical rating scales in research. Eligibility measures were administered only at Week 0, clinical outcome assessments (clinician and self-administered) were conducted at Weeks 0, 4 and 10 of the study, and feasibility and therapy process measures were administered at Weeks 4 and 10 (except where noted).

Eligibility Measures

Hamilton Rating Scale for Depression-17 Item (HAMD; Hamilton, 1960): The HAMD (a widely used measure of depressive symptom severity) was used to establish eligibility for study inclusion. HAMD scores range from 0–50, with higher scores indicating more severe

depressive symptoms. A score of 10 or more on the clinician-rated HAMD was required for inclusion. Internal consistency estimates for the HAMD across studies range from 0.46 to 0.97, and retest reliability across studies ranges from 0.81 to 0.98 (Bagby, Ryder, Schuller, & Marshall, 2004). Cronbach's alpha in this sample was 0.53. Most studies have shown adequate convergent validity (Bagby et al., 2004), and the HAMD also demonstrated good discriminant validity for classifying and discriminating among diagnostic groups (Bagby et al., 2004).

Structured Clinical Interview for DSM-IV Research Version (SCID; First et al. 2002):

The SCID is a structured clinical interview to establish psychiatric diagnoses. The Screening, Mood Disorders, Psychotic Disorders, OCD, Eating Disorders, and Substance Use Disorders Modules of the SCID were administered at Week 0 to determine diagnostic exclusion criteria and to establish mood disorder diagnoses. The Mood module was readministered at Week 10 to assess remission status. Test-retest reliability for the SCID modules used in this study ranges from 0.35 to 0.91 depending on module and study (Zanarini et al., 2000; Zanarini & Frankenburg, 2001). The SCID has demonstrated superior validity over clinical interviews for establishing psychiatric diagnosis and is considered the gold-standard for diagnostic interviewing (Steiner, Tebes, Sledge, & Walker, 1995).

Demographic data: At Week 0, participants completed a demographic questionnaire.

Feasibility and Therapy Process Measures

System Usability Scale (SUS; Brooke, 1996): The SUS is a 10-item self-report measure used to evaluate the ease of using interfaces, software, and other technological devices. Users endorse 10 statements (e.g., "I found the system unnecessarily complex") using a 5-point Likert scale (1 = *strongly disagree*; 5 = *strongly agree*). The sum of the individual items (range 0–40) is multiplied by 2.5 to obtain the total score, ranging from 0 to 100, with higher scores indicating better usability. Test-retest reliability has not been computed in any research of which we are aware. Cronbach's alpha across samples has been estimated at 0.91. The average SUS score across many studies of computer systems is 70 (Bangor, Kortum, & Miller, 2008). The SUS has demonstrated evidence of discriminant validity for detecting different interfaces and changes in products, as well as concurrent validity with a single-item measure of user-friendliness (Bangor et al., 2008). Cronbach's alpha for the SUS in the current sample ranged from .75 to .87.

Acceptability of Self-Guided Treatment (AST; Cartreine et al., 2012): The AST is a measure used in our previous research on *ePST*TM to assess the acceptability of *ePST*TM as a stand-alone treatment for depression (Cartreine et al.). This measure includes 16 statements (e.g., "I would feel comfortable using this program without a clinician's supervision") that participants respond to on a 7-point Likert scale, (1 = *strongly disagree*; 7 = *strongly agree*). Individual item scores are averaged and range from 1 to 7, where 7 represents strong agreement that *ePST*TM is acceptable as a stand-alone treatment for depression. Cronbach's alpha for the AST in this sample was excellent, ranging from .93 to .95. Alpha in the previous study was 0.83 (Cartreine et al.). Cartreine et al. did not present test-retest or validity data for the scale.

Credibility Questionnaire (CQ): The CQ is a 9-item, study-designed measure to assess the intervention's credibility. Items were drawn both from measures of computer program and website credibility (Fogg et al., 2001) and measures of the credibility of psychotherapy interventions (Borkovec & Nau, 1972). It also includes 2 items used in research on the acceptability of live PST (Thornett & Mynors-Wallis, 2002) (e.g., "How much do you trust the program to help you?"). Participants responded using a 10-point Likert scale (1 = *not at all*; 10 = *completely*). Individual item scores are averaged and range from 1–10, where 10 represents the intervention being completely credible to the user. Cronbach's alpha for the CQ in this sample was excellent, ranging from .95 to .97.

The Agnew Relationship Measure (ARM; Agnew-Davies, Stiles, Hardy, Barkham, & Shapiro, 1998): The ARM is a 28-item measure of the therapist-client working alliance in psychotherapy, designed to tap therapeutic bond, goal agreement, social influence, interpersonal initiative, and power aspects of the therapeutic alliance, without including items that might be early markers of outcome. We selected the ARM because it has been used in alliance research in CB treatment for depression (e.g., Ormrod et al., 2010; Reynolds et al., 2006). The ARM (client form) yields 5 subscales assessing the degree to which the client takes initiative for treatment and feels openness toward, confidence in, and partnership and an affective bond with the counselor. For the present study, items were adapted to reflect the computer context (e.g., "I feel friendly toward the therapist" became "I feel friendly toward the computerized therapist"). Items are rated on a 7-point Likert-type scale (1 = *strongly disagree*; 7 = *strongly agree*). Individual item scores are averaged and range from 1–7. Higher scores represent a stronger alliance with the therapist.

In previous research the ARM has demonstrated good internal consistency reliability (α N = 70) for all subscales except the Client Initiative subscale (α of 0.55 for client versions; Agnew-Davies et al., 1998). Test-retest reliability is impractical to obtain with the measure because it rates the quality of a psychotherapy session. Convergent validity has been demonstrated by high correlations with other measures of the therapeutic alliance and predictive validity has been demonstrated via positive correlations with psychotherapy outcome (Agnew-Davies et al.; Stiles et al., 2002). Previous research on the alliance in face-to-face psychotherapy has suggested that the alliance is generally stable over time (Martin et al., 2000), but that early assessments (i.e., after the 3-5th session of therapy) are considered most predictive of outcome and therapy dropout (Horvath, Del Re, Flückiger, & Symonds, 2011). Cronbach's alphas in our sample were generally acceptable or good. The Openness subscale had alphas ranging from .56 to .66. All other subscales had alphas above .7 at all assessments, except the client initiative subscale, which had some negative item intercorrelations both at Week 4 and Week 10.² We omitted the client initiative subscale from our analyses.

²The Client Initiative subscale had relatively low reliability even in its norming sample with live psychotherapy patients (alpha = .55 in Agnew-Davies et al., 1998), and has had unacceptably low reliability in other research on CB therapies (e.g., alpha = .26 in Ormrod et al., 2010), so our difficulties with this scale may be reflective of problems with this scale in general or when used with CB therapies. There was no single problematic item on the scale which could be omitted to improve scale reliability.

Clinical Outcome Measures

Hopkins Symptom Checklist 20–Item Depression Scale (HSCL; Katon et al., 1995): The HSCL is a self-report measure of depressive symptoms, and was used as a primary outcome measure in this study. This 20-item Depression scale is adapted from the 90-item HSCL (Derogatis, Lipman, Rickels, Uhlenbuth, & Covi, 1974; Lipman, Covi, & Shapiro, 1979). The HSCL has been a primary outcome measure in previous PST studies (e.g., Areal et al., 2008; Oxman et al., 2008), thus allowing direct comparisons between studies. The HSCL has shown good convergent and criterion-related validity, correlating with other measures of depression and demonstrating responsiveness to depression treatment (Johns et al., 2012; Williams et al., 2004). Items are rated on a 5-point scale (0–4) ranking how much the symptom has been experienced during the past week. Scale scores are determined by dividing the sum of the items by the total number of items; scores range from 0 to 4; higher scores indicate greater depression severity. Cronbach's alpha was .84–.94 over the course of the study, which is consistent with internal consistency estimates provided in other research (e.g., Johns et al., 2012; Williams et al., 2000).

Montgomery–Asberg Depression Rating Scale (MADRS; Montgomery & Asberg, 1979): We used the 10-item, clinician-rated MADRS as a second outcome measure. MADRS items are rated on a 0–6 severity scale, resulting in a total score range of 0–60; higher scores indicate greater depressive symptom severity. The MADRS has good test-retest reliability (Williams & Kobak, 2008) and convergent and discriminant validity (Brooks & Kutcher, 2001). It is sensitive to treatment change (Davidson, Turnbull, Strickland, & Miller, 1986; Montgomery & Asberg, 1979). A Cronbach's alpha of .40–.74 was observed over the course of this study.

Medical Outcomes Short Form-36 (SF-36; Ware & Kosinski, 2001): The SF-36 is a 36-item multidimensional measure of functioning (Ware & Kosinski, 2001), used here to measure functional impairment and quality of life. Total scores range from 0 to 100; high scores indicate better functioning. The SF-36 is reliable, with generally high internal consistency and test-retest reliability across studies (Ware, 2000). In terms of convergent and discriminant validity, the SF-36 has been correlated with a broad range of both mental and physical health outcomes and has generally shown expected correlations with such measures (Ware, 2000).

Patient Health Questionnaire-9 Item Depression Scale (PHQ-9; Kroenke et al., 2001): The PHQ-9 is the 9-item self-report depression screening module of the Patient Health Questionnaire. Participants rate 9 items assessing DSM-IV depression symptoms on a 0–3 scale with anchors ranging from “not at all” to “nearly every day” to endorse symptoms they've experienced in the past 2 weeks. Scores on the PHQ-9 range from 0–27; (0–4 = *no depression*; 5–9 = *mild depression*; 10–14 = *moderate depression*; 15–19 = *moderate-severe depression*; 20 or greater = *severe depression*). Participants completed the PHQ-9 at Week 1 and at each ePST™ session (6 sessions). The PHQ-9 has demonstrated strong internal consistency and test-retest reliability (Kroenke et al., 2001). In terms of criterion-related validity, PHQ-9 scores and severity cutoffs accurately categorize patients with no

depression, major depressive disorder, and a range of subclinical depressive symptoms (Kroenke et al.).

Comparison Samples for Exploratory Analyses—To put our results in context, we compared our findings with findings reported in published research on computerized and live psychotherapy (see Table 1). For the computerized therapy comparison we chose *Beating the Blues (BtB)*, a nine-session cognitive-behavioral depression treatment program (Ormrod et al., 2010). We chose *BtB* because (a) it is widely used and researched (Cavanagh et al., 2006; Learmonth, Trosh, Rai, Sewell, & Cavanagh, 2008; Proudfoot et al., 2003); (b) therapeutic alliance between *BtB* and its users has been investigated, providing a basis for comparison (Ormrod et al.); and (c) it has both key similarities to and differences from *ePST™*. Like *ePST™*, *BtB* uses a CBT approach; is intended as a stand-alone depression treatment with no mental health professional support; has sessions of comparable duration; provides customized homework assignments; and uses audio, video, and interactive text elements. Unlike *ePST™*, *BtB* has no virtual therapist. Users are shown video case studies of other patients, rather than receive individualized feedback about their own problem-solving efforts. Interactive and personalized elements are limited to text (e.g., checking or typing in pleasurable activities).

We compared our results to those of Ormrod et al. (2010) who used the ARM to investigate therapeutic alliance between users and *BtB*. Ormrod et al. included 23 participants (16 of whom completed the intervention) with a diagnosis of depression recruited from an adult mental health service in the U.K. Their sample was 52% female and roughly comparable to ours in age (mean age = 47, $SD = 11$, $N = 16$). Whereas our study used a Self-Administered Structured (SAS) intervention, their study employed a Predominantly Self-Help (PSH) intervention. Contact with a mental health professional during our study was limited to study assessments. In the Ormrod et al. study, participants were greeted at all sessions by a mental health professional, who reviewed their summary worksheets at the end of the sessions and offered an opportunity to ask questions or share concerns. In addition, following the intervention participants were offered a “review session” with a therapist to plan for any additional treatment needed. The *BtB* intervention ran for 9 consecutive weeks and the ARM was assessed weekly. Ormrod et al. report the mean values for the ARM subscales collected over all 9 weeks. We compared our Week 4 and Week 10 data to this value. Like Ormrod et al. we ran these comparisons using only our “completer” sample ($N = 23$).

For the live therapy comparison, we chose the Second Sheffield Psychotherapy Project, a randomized controlled trial of brief manualized psychodynamic and cognitive behavioral psychotherapies as treatments for depression. This comparison sample was chosen because it was the sample on which the ARM was developed (Agnew-Davies et al., 1998). Participants were treated with either 8 or 16 sessions of CBT or psychodynamic psychotherapy by clinical psychologists trained in both treatments. The sample was 51% female, and somewhat younger than our sample (mean age = 40, range 23–60). As *ePST™* is a brief (6 session) CBT intervention, we compared our findings to the Sheffield brief (8-session) CBT condition, using data obtained from the researchers ($N = 72–74$). Like the Ormrod et al. (2010) *BtB* study, the Sheffield Project administered the ARM after every session. We conducted our comparisons against the mean value for the first 4 assessments

(i.e., Week 4) and the last 4 assessments (i.e., week 8) for our *ePST*TM data from Week 4 and Week 10, respectively, using our completer sample ($N = 23$).

For depression outcomes, we compared our participants to those in a previous randomized controlled trial conducted by our team which compared live PST to usual primary care treatment of minor depression (Oxman et al., 2008). In this previous trial, masters'-level clinicians delivered manualized PST-PC over 6 weekly or biweekly sessions. Participants randomized to usual care received a single appointment with their primary care provider, who then managed their care as they thought appropriate. Providers could suggest watchful waiting, prescribe antidepressants, provide brief supportive counseling, and/or refer to specialty mental health. Less than 30% of usual-care participants received a prescription for antidepressants. Similarly, only 6% of participants in usual care reported having one or more appointments with an outside mental health professional, whereas 81% of participants had at least one depression-related visit with their primary care provider.

As in our current study, all interventions took place over 9 weeks, and study assessments used the same measures at the same time points (Weeks 0, 4 and 10). Participants in the prior trial were similar geographically and demographically to our participants (see Results section for details). We chose this study for a comparison to live PST because we wanted to create the strongest possible comparison between live PST and *ePST*TM. The Oxman et al. (2008) study had strong similarities to ours in research team, design, measures, and participants. In addition, because this study was conducted by our research team, we had access to all the data available from the study. We also compared our data to those reported by Mynors-Wallis et al. (1995), which included a population with a depression severity level more similar to our sample.

Statistical Analysis—Descriptive data are reported to address Hypotheses 1 and 2 (that participants would rate *ePST*TM highly credible, acceptable and usable, and that the program would create a strong therapeutic alliance with participants). To evaluate the therapeutic alliance developed, we performed independent-samples t-tests comparing our *ePST*TM participants to the published comparison samples described above (Agnew-Davies et al., 1998; Ormrod et al., 2010). To examine Hypothesis 3, that depression would show clinically and statistically significant improvement over time, and compare depression outcomes in *ePST*TM with those for therapist-directed PST, we conducted a doubly multivariate repeated measures analysis of variance with comparisons over time between our sample and participants from a previous study by our group (Oxman et al., 2008). A separate doubly multivariate analysis was applied to our measures of health functioning (the three subscales of the SF-36). SPSS software was used for all statistical analyses. All multivariate F-values for these analyses are based on Wilks-Lambda estimates. Effect sizes are partial eta-squares as reported by SPSS. No protections for multiple tests were applied when differences were not predicted between compared groups.

Results

ATTRITION

Figure 1 illustrates the flow of participants through the study. Of the 29 enrolled participants, 23 completed an adequate dose of *ePST*TM (4+ sessions; 87%). This represents a similar attrition rate to that observed in live PST (79% in Oxman et al., 2008). The completion rate for the entire program was also high, with 21 participants completing all six treatment sessions. Completers were compared with noncompleters ($n = 6$) on baseline demographic and clinical characteristics. Noncompleters were significantly more depressed at baseline than completers on the MADRS, $t(27) = -2.43, p = .02$, but rated themselves as having significantly less emotional impairment than completers on the role emotional scale, $t(27) = -2.32, p = .03$. There were no other significant differences between completers and non-completers on baseline measures.

At the 4-week assessment, noncompleters found *ePST*TM significantly less usable than completers, $t(24) = 2.44, p = .02$. There were no significant differences in acceptability or credibility between completers and non-completers at Week 4. At Week 10, noncompleters rated the system as significantly less acceptable, $t(25) = 2.17, p = .04$, usable, $t(25) = 2.72, p = .01$, and credible, $t(25) = 2.49, p = .02$, than completers. Noncompleters and completers did not differ on any of the therapeutic alliance (ARM) subscales at Weeks 4 or 10 (N s for these analyses ranged from 26 to 28 because of missing data from some noncompleters who were unavailable for follow-up assessment). Of note, two of the noncompleters discontinued the study specifically because they were advised by the on-screen therapist to seek care from their doctor or a mental health professional based on the severity of their symptoms as assessed within the program. They followed this advice and discontinued *ePST*TM.

USABILITY, ACCEPTABILITY, CREDIBILITY

*ePST*TM participants rated system usability, acceptability, and credibility at Weeks 4 and 10. Participants ($N = 26$) found the program significantly more credible (CQ) at Week 10 (mean = 7.5, $SD = 2$) compared to Week 4 (mean = 6.7, $SD = 1.9$), $t(24) = -3.25, p = .003$ (1–10 scale; 1 = *not at all credible*, 4 = *a little*, 7 = *a lot*, and 10 = *completely credible*). Means and standard deviations for study measures at all time points are presented in Table 2.

The mean score for acceptability of *ePST*TM as a stand-alone treatment for depression (AST) remained stable from Week 4 (mean = 5.29, $SD = 1.1$) to Week 10 (mean = 5.46, $SD = 1.3$), $t(24) = -1.38, ns$ (1–7 scale; 1 = *strongly disagree*, 4 = midpoint, 7 = *strongly agree*).

*ePST*TM participants found the program significantly more usable (SUS) than the average system as reported by Bangor et al. (2008) in their meta-analysis (mean = 70) at Week 4, $N = 27, t(25) = 3.27, p = .003$, and Week 10, $N = 28, t(26) = 2.8, p = .01$. The mean score at Week 10 was 79.4 ($SD = 17.3$) on the 1–100 scale. This score reflects system usability in the good-to-excellent range (Bangor et al., 2008).

THERAPEUTIC ALLIANCE

Participants rated therapeutic alliance (ARM) with the *ePST*TM program at 4 and 10 weeks. The only significant changes on the ARM subscales over time were that the confidence

*ePST*TM participants ($N = 26$) reported in the program and the virtual therapist increased significantly from 4 to 10 weeks, $t(24) = -2.79, p = .01$. Table 3 displays means and standard deviations for ARM subscales at weeks 4 and 10 for *ePST*TM. Table 3 also displays these values for both live CBT (Agnew-Davies et al., 1998) at 4 and 8 weeks, and another computer-based therapy, *BtB* (Ormrod et al., 2010), at 9 weeks.

At Week 10, *ePST*TM was superior to *BtB* on all ARM subscales: Bond, $t(37) = 5.01, p < 0.001, d = 1.57$; Partnership, $t(37) = 2.26, p = 0.03, d = 0.70$; Confidence, $t(37) = 4.27, p < 0.001, d = 1.36$; Openness, $t(37) = 4.33, p < 0.001, d = 1.38$). Similarly, at Week 4, *ePST*TM was superior to *BtB* for all subscales: Bond, $t(37) = 4.68, p < 0.001, d = 1.48$; Confidence, $t(37) = 2.8, p = 0.01, d = 0.89$; Openness, $t(37) = 2.68, p = 0.01, d = 0.84$, except Partnership, $t(37) = 1.30, p = 0.2, d = 0.40$. There were no significant differences between *ePST*TM and Live CBT assessments on any ARM subscale except for Partnership. Live CBT was superior to *ePST*TM for Partnership at both Week 4, $t(95) = -2.31, p = 0.02, d = -0.59$, and Weeks 8/10, $t(169) = -2.06, p = 0.04, d = -0.50$.

Live CBT significantly outperformed *BtB* on all subscales at both time points: Week 4 Bond, $t(87) = -4.56, p < 0.001, d = -1.03$; Week 4 Partnership, $t(88) = -5.6, p < 0.001, d = -1.33$; Week 4 Confidence, $t(86) = -4.68, p < 0.001, d = -1.18$; Week 4 Openness, $t(87) = -3.35, p = 0.001, d = -0.78$; Week 8/10 Bond, $t(159) = -5.85, p < 0.001, d = -1.19$; Week 8/10 Partnership, $t(162) = -6.77, p < 0.001, d = -1.49$; Week 8/10 Confidence, $t(159) = -5.46, p < 0.001, d = -1.25$; Week 8/10 Openness, $t(163) = -5.13, p < 0.001, d = -1.09$.

DEPRESSION HSCL AND MADRS

Comparison With Oxman et al. (2008)—We hypothesized that depressive symptoms would show significant improvement over time in our sample. To provide context for these findings we also compared improvement with *ePST*TM to that obtained with live PST and the control condition (usual care) in a previous minor depression study (Oxman et al., 2008). Table 4 displays baseline characteristics comparing our sample with these two comparison conditions. At baseline, the *ePST*TM group was more severely depressed and had poorer social functioning compared with both minor depression groups. Oxman et al. specifically excluded individuals with a major depressive disorder or dysthymic disorder diagnosis, whereas most of our sample met criteria for major depression or dysthymia. Thus, the *ePST*TM participants were more severely depressed. Our sample was also significantly more educated than the comparison sample. There were no other significant differences between the samples at baseline.

To determine whether the *ePST*TM and live PST groups responded similarly to the interventions and to investigate improvement over time in the *ePST*TM group, a doubly multivariate repeated measures analysis of variance was conducted that involved one between-subjects factor with three levels (Treatment: *ePST*TM, $N = 25$, PST, $N = 54$; Usual Care, $N = 47$), a repeated factor with three levels (Time: T1, T2, T3), and two measures of depression (HSCL, MADRS). This analysis revealed a significant main effect of Time, $F(4,490) = 70.27, p < .001, \eta^2 = .30$, such that all participants decreased in depressive symptoms over time for both measures. There was also a main effect of Treatment, $F(4,244)$

$= 2.528, p < .05, \eta^2 = .04$, such that the *ePST*TM group was higher in depression than either the live PST or Usual Care group at all time points. (Figures 2A and 2B).

The Time \times Treatment interaction was nonsignificant, $F(8, 490) = 1.685, p < .10, \eta^2 = .03$. In the previous study by Oxman et al. (2008) a significant difference was found between live PST and Usual Care between Time 1 and Time 2; differences between live PST and Usual Care were nonsignificant between Times 2 and 3. If *ePST*TM mimics live PST in its effects, then the slope of improvement should be similar for these two conditions from Time 1 to Time 2 and different from the slope of improvement in the Usual Care condition. This is supported by several subsidiary analyses. A planned contrast reflecting the Time \times Treatment interaction for the phase Time 1 – Time 2 is significant for the MADRS, $F(2, 123) = 3.34, p = .04, \eta^2 = .05$; for the HSCL, $F(2, 123) = 3.07, p = .05, \eta^2 = .05$. These findings reflect that the treatment conditions differ in the rate of improvement in depression over the early phase of treatment. Restricting analyses to the Time 1 – Time 2 periods and analyzing treatment conditions separately, the Time main effect for the HSCL is statistically significant and the effect size large among PST participants, $F(1, 59) = 45.21, p < .001, \eta^2 = .43$, and *ePST*TM participants, $F(1, 25) = 18.20, p < .001, \eta^2 = .42$, but not among usual-care participants, $F(1, 52) = 2.62, p = .11, \eta^2 = .05$. The pattern is similar for the MADRS, although in this case the Time main effect is significant for PST, $F(1, 61) = 97.71, p < .001, \eta^2 = .62$, *ePST*TM, $F(1, 25) = 40.45, p < .001, \eta^2 = .62$, and usual-care participants, $F(1, 54) = 17.22, p < .001, \eta^2 = .24$.

In contrast to these observed effects, the Time \times Treatment interaction for the phase Time 2 – Time 3 is not significant for either the HSCL, $F(2, 123) = 1.54, p > .20, \eta^2 = .02$, or the MADRS, $F(2, 123) = 1.26, p > .20, \eta^2 = .02$. Similarly, the Time \times Treatment interaction for the phase Time 1 – Time 3 is not significant for either the HSCL, $F(2, 123) = 1.59, p > .20, \eta^2 = .03$, or the MADRS, $F(2, 123) = .67, p > .20, \eta^2 = .01$. These results are consistent with the conclusion that the treatment conditions differ in rate of improvement in depression in the early but not later phase of treatment, with larger effects associated with PST and *ePST*TM than usual care.

Comparison With Mynors-Wallis et al. (1995)—A second comparison involved live PST and placebo conditions from a previous study of depression by Mynors-Wallis et al. (1995). Although both studies used the HAMD to assess depression at baseline, for followup assessments we used the clinician-rated MADRS and self-rated HSCL, whereas Mynors-Wallis et al. used the clinician-rated HAMD and the self-rated Beck Depression Inventory (Beck, Ward, & Mendelson, 1962). Furthermore, follow-up assessments in the two studies occurred at different time points. Despite these differences, some rough comparisons can be made. If the same inclusion criteria used by Mynors-Wallis et al. (participants with initial HAMD scores greater than or equal to 13) are applied to our sample, the two samples are comparable in initial depression levels (Mynors-Wallis mean HAMD = 18.4; our mean HAMD = 18.0). Using a criteria of 7 or less on the HAMD to define a participant as asymptomatic (see Frank et al., 1991), Mynors-Wallis reported that 27% (8 out of 30) of participants in their placebo condition and 60% (18 out of 30) of participants in their problem solving condition were asymptomatic at 12 weeks—a statistically significant difference, $\chi^2(1) = 6.79, p = .01, \phi = .34$. According to Zimmerman, Posternak, and

Chelminski (2004), a score of 10 or less on the MADRS is comparably asymptomatic (similar comparison criteria do not exist for the HSCL). Using this criterion, we observe that 58% (14 out of 24) of participants in our *ePST*TM condition were asymptomatic at 10 weeks. Admittedly, comparing these outcomes involves using different scales at different time points to determine participants' asymptomatic status. Nonetheless, it is encouraging that the difference in proportion remitted across these three groups (PST, placebo, and *ePST*TM) is statistically significant, $\chi^2(2) = 8.23, p = .02, \phi = .31$. Furthermore, a direct comparison of *ePST*TM and PST in proportion remitted is not significant, $\chi^2(1) = .02, p = .90, \phi = .02$, whereas a greater proportion of participants receiving *ePST*TM are remitted compared to those receiving a placebo, $\chi^2(1) = 5.54, p = .02, \phi = .32$.

DEPRESSION: PHQ-9

Participants completed the PHQ-9 at prescreen and at the beginning of each *ePST*TM session. We conducted a repeated measures ANOVA to test the hypothesis that PHQ-9 scores would improve over time. The main effect of time was significant, $N = 20$ (due to incomplete data), $F(6, 14) = 31.558, p < .001, \eta^2 = .931$. Figure 3 displays mean scores on the PHQ-9 over time. At pre-screen, the mean PHQ-9 score was 12.63, representing moderate depression. By Session 6, the mean score was 4.9, in the nondepressed range.

HEALTH FUNCTIONING

To test the hypothesis that *ePST*TM participants' outcomes for health functioning would reflect improvement over time and match the improvements made in the Oxman et al. (2008) live PST comparison sample, a doubly multivariate repeated measure analysis of variance was conducted for functioning as measured by three subscales of the SF-36 (Physical Functioning, Social Functioning, and Role Emotional Functioning) at two time points (Time 1 and Time 3). (Because this measure was only obtained in the Oxman et al. comparison sample at baseline and follow-up, we did not include the midpoint assessment in our analyses.) This analysis revealed a significant main effect of Time, $F(3, 126) = 739.727, p < .001, \eta^2 = .36$. Univariate analyses revealed that all participants improved in functioning over time on all three measures. In addition, there was a main effect of treatment, $F(6, 252) = 3.359, p < .01, \eta^2 = .07$, such that the *ePST*TM group ($N = 27$) was generally lower in functioning than either the PST ($N = 54$) or usual-care group ($N = 50$). Univariate analyses revealed that this difference was only significant for social functioning. Oneway ANOVAs revealed that this difference was significant at both baseline, $F(2, 167) = 7.758, p = .001$, and follow-up, $F(2, 128) = 3.359, p = .038$. Neither the multivariate interaction of Time \times Treatment, $F(6, 252) = 1.577, p = .15, \eta^2 = .04$, nor any of the univariate interactions approached significance (all $p > .15$).

CLINICAL SIGNIFICANCE

The clinical significance of treatment effects was assessed in a number of ways. Following the recommendations of Jacobson (Jacobson, Roberts, Berns, & McGlinchey, 1999; Jacobson & Truax, 1991), we used results from the development of the Hopkins Symptom Checklist, specifically norms derived from the normal control subsample (Derogatis et al., 1974) to define the mean and standard deviation of the HSCL depression scale in the normal

population. Given that our initial sample distribution overlapped the normal comparison group, a clinically significant change in functioning subsequent to *ePST*TM was defined as an outcome HSCL score (as measured at Week 10) closer to the mean of the norm group than to the mean of the initial sample (Jacobson & Truax, 1991). As a consequence, the criterion for clinically significant change was an HSCL depression score of 1.38. Of participants with complete data ($N = 26$), 73% achieved this criterion. Jacobson and Truax also recommended using the Reliable Change Index to assess the reliability of change within the sample. In the present sample, a reduction on the HSCL of 0.33 units constitutes reliable change. Relative to initial baselines, 95% of participants with complete data who achieved clinically significant improvement also met this criterion for reliable change. This is roughly comparable to our comparison sample (Oxman et al., 2008) in which 83% of live PST participants achieved clinically significant change on the HSCL by Week 9, of whom 78% met the criteria for reliable change. In their usual-care group, 84% achieved clinically significant change, of whom 67% met the criteria for reliable change.

Psychiatric diagnoses also changed. At baseline, 83% of *ePST*TM participants met criteria for either or both major depressive disorder or dysthymic disorder.³ At Week 10, only 18% of participants met criteria for either or both major depression or dysthymia.

Discussion

USABILITY, CREDIBILITY, ACCEPTABILITY, AND THERAPEUTIC ALLIANCE

*ePST*TM participants rated the program as acceptable, usable and credible, and reported a strong therapeutic alliance with the program. *ePST*TM was designed to engage the user in a fashion that mimics face-to-face psychotherapy. The program provides personalized feedback from a virtual therapist who appears on screen and speaks directly and encouragingly to the user. Extensive use of audio and video technology enhances the simulation, and hundreds of branching audio, video, and graphic elements allow the therapist to respond flexibly to users' changing levels of depression, their struggles or successes implementing therapeutic homework, and their understanding of the therapy. These elements help to create the experience of having a caring therapist present. In addition, users benefit from receiving PST with exact treatment fidelity from an expert therapist, who conveys the "non-specific" characteristics of effective therapists, such as warmth, genuineness, compassion, and support when the patient experiences setbacks.

The *ePST*TM program was feasible and acceptable. The number of noncompleter participants in our sample was small ($n = 6$), as was the number of completers ($n = 23$) thus making comparisons between the groups difficult. Compared to completers, noncompleters were more depressed, more likely to be male, and less likely to report emotional impairment (despite being more depressed). They also found the program less acceptable, usable, and credible. Further research is indicated to determine whether noncompleters in *ePST*TM differ

³Three participants did not have complete SCID interview data at baseline to derive an MDD diagnosis and did not meet criteria for Dysthymic Disorder; 2 participants did not meet full criteria for MDD or dysthymic disorder despite having significant depressive symptoms.

from participants who drop out of live therapy, and to understand what leads to poor engagement with the program for participants who drop out.

Participants who completed treatment rated *ePST*TM as highly usable (mean score in the good-to-excellent range), credible (mean score in the “a lot” of credibility range) and logical (median scores on “logical” and “recommend it to a friend” were comparable to those obtained in research on live PST; Thornett & Mynors-Wallis, 2002). On average, participants felt that *ePST*TM was acceptable as a stand-alone treatment for depression (indicating that they felt they did not need a clinician’s input).

This study provides evidence that the program was successful in creating a therapeutic alliance comparable to live CBT (Agnew-Davies et al., 1998). Within the limits of our analyses, we found no significant differences in therapeutic bond, confidence in the therapist and program, or willingness to be open in discussing concerns. Participants in live CBT did report a stronger partnership with their therapist. Agnew-Davies et al. found that older participants reported greater alliance than did younger participants. Since the average age of our sample was 53 years old, this may account to some extent for the strong therapeutic alliance. Additional research with a broader age group is warranted to determine whether younger users will develop an alliance with *ePST*TM.

The *ePST*TM participants’ alliance scores were consistently higher than those obtained in research on another computer based program, *BtB* (Ormrod et al., 2010). At both mid- and posttreatment assessments, *ePST*TM was superior to *BtB* on forming a therapeutic bond, feeling a sense of partnership with the program, confidence that the program would help them, and feeling comfortable openly discussing their concerns with the program. These findings suggest that the goal of creating a strong sense of alliance with the *ePST*TM program was accomplished. Further research is indicated to identify the specific mechanisms whereby this occurs.

DEPRESSION AND HEALTH FUNCTIONING

During the study, *ePST*TM depression and health-related functioning improved significantly. The MADRS, HSCL, and PHQ-9 scores all improved significantly over time. These improvements were clinically meaningful as well as statistically significant: 73% of the study completers (26/29) met the criteria for a clinically meaningful treatment effect, of whom 95% met the criteria for reliable change. The proportion of participants meeting DSM-IV diagnostic criteria for major depression and/or dysthymic disorder fell from 83% at baseline to 18% following treatment with *ePST*TM. Furthermore, participants significantly improved over time in social functioning and role functioning related to both emotional and physical symptoms.

COMPARISONS OF EPSTTM TO PRIOR RESEARCH WITH PROBLEM SOLVING TREATMENT FOR DEPRESSION

To overcome some of the limitations of a single-arm trial, we directly compared our findings to those from one of our previous studies of PST-PC for minor depression with live therapists (Oxman et al., 2008). Although the studies had similar designs, the current study

did not limit eligibility solely to minor depression. As a result, depression severity was significantly higher at baseline in the *ePST*TM group and remained significantly higher across all assessment points. Otherwise, the results from the current study were comparable to the original Oxman et al. study. Both *ePST*TM and live PST groups improved significantly faster than the usual-care control condition between baseline and the 4-week assessment points, although all groups improved significantly over time. Nonetheless, because depression severity for the *ePST*TM group was significantly higher at baseline, we cannot rule out regression to the mean as an explanation for the improvement. In part to address this issue, we also compared *ePST*TM to the published findings of a randomized placebo controlled trial of live PST for moderately severe depression (Mynors-Wallis et al., 1995). Using established scale-based criteria for remission of depression, we found that *ePST*TM was comparable to live PST and superior to the placebo condition.

LIMITATIONS

There are several limitations to this study due to the single-arm design. We did not have an experimental control group; therefore, improvements we observed in depression outcomes may have been due to the passage of time, spontaneous remission in depression, regression to the mean, or other factors apart from the *ePST*TM program. Where possible, we attempted to address this limitation by comparing our findings to relevant findings in the literature to try to estimate how *ePST*TM might compare to treatments evaluated in randomized controlled trials. Nonetheless, comparisons to the Oxman et al. (2008) study are still problematic given that our sample was significantly more depressed at baseline and remained more depressed throughout the study. In addition, clinician raters for interview measures (MADRS, SCID) were not blind to treatment assignment or study hypotheses and this could have biased their ratings. In the case of the MADRS, low or unknown reliability of the measure could also have affected the ratings obtained. We included self-report measures of depression, with good reliability, to address this limitation. Results on the self-report measures were similar to those obtained on the clinician-rated measures.

A second limitation is our small sample size, which limited our ability to detect group differences in the comparisons with other research. A small sample is appropriate for a feasibility study like this one, but our conclusions should be considered tentative, to be confirmed with replication in larger samples.

A third limitation involves the comparison of our results to those from other published studies. Although we chose comparison studies that used similar methods and participants, a variety of known and unknown differences between our study and these comparison studies may have influenced our findings. For the depression symptom comparisons, this concern is ameliorated somewhat by the fact that one comparison study (Oxman et al., 2008) was done by our research team, at our institution, and using many of the same measures and overall study design. For the therapeutic alliance and credibility comparisons, differences between our study and comparison studies are more difficult to quantify. Randomized controlled trials comparing different computer-based therapies, to the best of our knowledge, have not been conducted, and perhaps are not likely to be conducted due to proprietary commercial

interests. We chose to include these analyses in an effort to elucidate possible differences in these important areas. Conclusions should be drawn with caution.

Fourth, we did not include a long-term follow-up assessment in this trial, and have no data on the long-term effectiveness of the intervention. Because relapse is a significant problem in depressive disorders, the long-term outcome of *ePST*TM should be assessed in future research.

Finally, in this study, *ePST*TM was delivered as a self-administered, structured (SAS) intervention. Participants had no contact with a therapist during their study participation except for study assessments and (if suspected of being suicidal) safety assessments. They were, however, required to make an appointment to use the program and check in with a (nonprofessional) study coordinator who set up the computer and was available in another room if questions arose. Therefore, our findings cannot be generalized to a fully stand-alone computer-based treatment. Testing *ePST*TM as a fully stand-alone treatment is an important direction for future research.

CONCLUSIONS

Participants rated the *ePST*TM program as highly usable, acceptable, and credible, and few participants dropped out. Participants reported a therapeutic alliance that was strong and comparable to that observed in live therapy. Participants who completed *ePST*TM improved significantly in depressive symptoms, to a degree similar to that observed in live treatment. Further research, particularly a full-scale randomized controlled trial, is indicated to evaluate the effectiveness of *ePST*TM as a stand-alone treatment and elucidate the mechanisms by which CB treatments can facilitate a therapeutic alliance.

Acknowledgments

James A. Cartreine was primarily responsible for the design and development of *ePST*TM. We would like to thank Angelica Strassberger for her invaluable assistance conducting this project, and Cynthia Stewart for coordinating the use of the library space for administering the *ePST*TM intervention. We thank Bill Stiles from Miami University in Ohio for sharing the ARM data from the Second Sheffield Psychotherapy Project. We appreciate the National Space Biomedical Research Institute for supporting the development of the *ePST*TM program through NASA NCC 9-58.

Inquiries about *ePST*TM should be directed to info@cognitivebehavioraltechnologies.com.

Disclosure Statement

This research was funded by a Dartmouth Center for Clinical and Translational Research SYNERGY Pilot Project Grant awarded to Mark Hegel. The funding source had no role in study design, execution, write up, or publication. Ownership of the *ePST*TM program was assigned to Cognitive Behavioral Technologies, LLC. Initially, Drs. Hegel and Buckey were partners in this company. Prior to conducting the study, but less than 3 years prior, Dr. Buckey divested his interests in this company. During the time the study was conducted, Dr. Hegel was a principal partner, but has since divested his interests in the company. The remaining authors have no conflicts of interest to disclose. The following conflict of interest management plan was employed. Dr. Berman oversaw the conduct of the study, including participant recruitment, intervention, and assessments. Drs. Berman and Hull conducted the data analyses. *ePST*TM is a trademark of Cognitive Behavioral Technologies, LLC.

References

- Agnew-Davies R, Stiles WB, Hardy GE, Barkham M, Shapiro DA. Alliance structure assessed by the Agnew Relationship Measure (ARM). *British Journal of Clinical Psychology*. 1998; 37(2):155–172. <http://dx.doi.org/10.1111/j.2044-8260.1998.tb01291.x>. [PubMed: 9631204]
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th ed. Author; Washington, DC: 1994.
- Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: A meta-analysis. *Cognitive Behaviour Therapy*. 2009; 38(4):196–205. <http://dx.doi.org/10.1080/16506070903318960>. [PubMed: 20183695]
- Andersson G, Paxling B, Wiwe M, Vernmark K, Felix CB, Lundborg L, Carlbring P. Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder. *Behaviour Research and Therapy*. 2012; 50:544–550. <http://dx.doi.org/10.1016/j.brat.2012.05.003>. [PubMed: 22728647]
- Arean P, Hegel M, Vannoy S, Fan M, Unutzer J. Effectiveness of problem-solving therapy for older, primary care patients with depression: Results from the IMPACT project. *The Gerontologist*. 2008; 48(3):311–323. <http://dx.doi.org/10.1093/geront/48.3.311>. [PubMed: 18591356]
- Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: Has the gold standard become a lead weight? *American Journal of Psychiatry*. 2004; 161(12):2163–2177. [PubMed: 15569884]
- Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*. 2008; 24(6):574–594. <http://dx.doi.org/10.1080/10447310802205776>.
- Barazzone N, Cavanagh K, Richards DA. Computerized cognitive behavioural therapy and the therapeutic alliance: A qualitative enquiry. *British Journal of Clinical Psychology*. 2012; 51(4):396–417. [PubMed: 23078210]
- Beck AG, Ward CH, Mendelson M. An inventory for measuring depression. *Archives of General Psychiatry*. 1962; 4:361–371.
- Bickmore T, Gruber A, Picard R. Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*. 2005; 59(1):21–30. [PubMed: 16198215]
- Borkovec TD, Nau SD. Credibility of analogue therapy rationales. *Journal of Behavior Therapy and Experimental Psychiatry*. 1972; 3(4):257–260. [http://dx.doi.org/10.1016/0005-7916\(72\)90045-6](http://dx.doi.org/10.1016/0005-7916(72)90045-6).
- Brooke, J. SUS: A “quick and dirty” usability scale.. In: Jordan, PW.; Thomas, B.; Weerdmeester, BA.; McClelland, IL., editors. *Usability evaluation in industry*. Taylor & Francis; London, UK: 1996. p. 189-194.
- Brooks SJ, Kutcher S. Diagnosis and measurement of adolescent depression: A review of commonly utilized instruments. *Journal of Child and Adolescent Psychopharmacology*. 2001; 11(4):341–376. [PubMed: 11838819]
- Carter JA, Buckey JC, Greenhalgh L, Holland AW, Hegel MT. An interactive media program for managing psychosocial problems on long-duration space-flights. *Aviation, Space, and Environmental Medicine*. 2005; 76:B213–B223. Retrieved from <http://search.proquest.com/docview/620781464?accountid=10422>.
- Cartreine JA, Locke S, Buckey JC, Hegel MT. Electronic Problem-solving Treatment (ePST™): Description and pilot study of an interactive media treatment for depression. *Journal of Medical Informatics Research*. 2012; 1(2):1–13. <http://www.researchprotocols.org/2012/2/e11/>.
- Cavanagh K, Shapiro DA, Van Den Berg S, Swain S, Barkham M, Proudfoot J. The effectiveness of computerized cognitive behavioural therapy in routine care. *British Journal of Clinical Psychology*. 2006; 45(4):499–514. <http://dx.doi.org/10.1348/014466505X84782>. [PubMed: 17076960]
- Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: Analysis of site usage and change in depression and anxiety scores. *Journal of Medical Internet Research*. 2002; 4(1):e3. <http://dx.doi.org/10.2196/jmir.8.4.e31/jmir.4.1.e3>. [PubMed: 11956035]

- Davidson J, Turnbull CD, Strickland R, Miller R. The Montgomery-Asberg Depression Scale: Reliability and validity. *Acta Psychiatrica Scandinavica*. 1986; 73(5):544–548. <http://dx.doi.org/10.1111/j.1600-0447.1986.tb02723.x>. [PubMed: 3751660]
- Derogatis LR, Lipman RS, Rickels K, Uhlenbuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioural Science*. 1974; 19:1–15.
- First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured Clinical Interview for DSM-IV-TR Axis I Disorders–Patient edition (with Psychotic Screen). New York State Research Institute; New York, NY: 2002. <http://dx.doi.org/10.1521/pedi.1995.9.2.92>
- Fogg, BJ.; Marshall, J.; Laraki, O.; Osipovich, A.; Varma, C.; Fang, N.; Treinen, M. What makes Web sites credible? A report on a large quantitative study.. CHI 2001 conference proceedings: Conference on Human Factors in Computing Systems; New York, NY: ACM Press; 2001. <http://dx.doi.org/10.1145/365024.365037>
- Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, Weissman MM. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence. *Archives of General Psychiatry*. 1991; 48:851–855. [PubMed: 1929776]
- Hamilton M. A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*. 1960; 23:56–62. <http://dx.doi.org/10.1136/jnnp.23.1.56>.
- Horvath, AO.; Del Re, AC.; Flückiger, C.; Symonds, D. Alliance in individual psychotherapy.. In: Norcross, JC., editor. *Psychotherapy relationships that work: Evidence-based responsiveness*. 2nd ed.. Oxford University Press; New York: 2011.
- Hirschfeld RMA, Keller MB, Panico S, Arons BS. The national depressive and manic-depressive association consensus statement on the undertreatment of depression. *Journal of the American Medical Association*. 1997; 277:333–340. <http://dx.doi.org/10.1001/jama.277.4.333>. [PubMed: 9002497]
- Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*. 1999; 67(3):300–307. <http://dx.doi.org/10.1037/0022-006X.67.3.300>. [PubMed: 10369050]
- Jacobson, NS.; Truax, P. Methodological issues & strategies in clinical research. *American Psychological Association*; Washington, DC: 1991. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research.; p. 631-648.<http://dx.doi.org/10.1037/10109-042>
- Johns SA, Kroenke K, Krebs EE, Theobald DE, Wu J, Tu W. Longitudinal comparison of three depression measures in adult cancer patients. *Journal of Pain and Symptom Management*. 2012; 45(1):71–82. [PubMed: 22921152]
- Kalenthaler E, Brazier J, De Nigris E, Tumur I, Ferriter, Michael, Beverley C, Sutcliffe PA. Computerised cognitive behaviour therapy for depression and anxiety update: A systematic review and economic evaluation. *Health Technology Assessment*. 2006; 10:1–186. <http://dx.doi.org/10.3310/hta10330>.
- Katon W, Von Korff M, Lin E, Walker E, Simon GE, Bush T, Russo J. Collaborative management to achieve treatment guidelines: Impact on depression in primary care. *Journal of the American Medical Association*. 1995; 273:1026–1031. <http://dx.doi.org/10.1001/jama.1995.03520370068039>. [PubMed: 7897786]
- Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Wang PS. The epidemiology of major depressive disorder: Results from the national comorbidity survey replication. *Journal of the American Medical Association*. 2003; 289(23):3095–3105. <http://dx.doi.org/10.1001/jama.289.23.3095>. [PubMed: 12813115]
- King R, Bambling M, Reid W, Thomas I. Telephone and online counselling for young people: A naturalistic comparison of session outcome, session impact and therapeutic alliance. *Counselling & Psychotherapy Research*. 2006; 6(3):109–115. <http://dx.doi.org/10.1080/14733140600874084>.
- Knaevelsrud C, Maercker A. Does the quality of the working alliance predict treatment outcome in online psychotherapy for traumatized patients? *Journal of Medical Internet Research*. 2006; 8:e31. <http://dx.doi.org/10.2196/jmir.8.4.e31>. [PubMed: 17213049]

- Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*. 2001; 16(9):606–613. <http://dx.doi.org/10.1046/j.1525-1497.2001.016009606.x>. [PubMed: 11556941]
- Learmonth D, Trosh J, Rai S, Sewell J, Cavanagh K. The role of computer-aided psychotherapy within an NHS CBT specialist service. *Counselling & Psychotherapy Research*. 2008; 8(2):117–123. <http://dx.doi.org/10.1080/14733140801976290>.
- Leibert T, Archer J Jr. An exploratory study of client perceptions of internet counseling and the therapeutic alliance. *Journal of Mental Health Counseling*. 2006; 28(1):69–83.
- Lipman RS, Covi L, Shapiro AK. The Hopkins Symptom Checklist (HSCL). *Journal of Affective Disorders*. 1979; 1(1):9–24. [http://dx.doi.org/10.1016/0165-0327\(79\)90021-1](http://dx.doi.org/10.1016/0165-0327(79)90021-1). [PubMed: 162184]
- Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*. 2000; 68(3): 438–450. <http://dx.doi.org/10.1037/0022-006X.68.3.438>. [PubMed: 10883561]
- Mohr DC, Siddique J, Ho J, Duffecy J, Jin L, Fokuo JK. Interest in behavioral and psychological treatments delivered face-to-face, by telephone, and by internet. *Annals of Behavioral Medicine*. 2010; 40(1):89–98. <http://dx.doi.org/10.1007/s12160-010-9203-7>. [PubMed: 20652466]
- Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*. 1979; 134(4):382–389. <http://dx.doi.org/10.1192/bjp.134.4.382>. [PubMed: 444788]
- Mynors-Wallis LM, Gath DH, Lloyd-Thomas AR, Tomlinson D. Randomised controlled trial comparing problem solving treatment with amitriptyline and placebo for major depression in primary care. *British Medical Journal*. 1995; 18(310):441–445. [PubMed: 7873952]
- Mynors-Wallis LM, Gath DH, Day A, Baker F. Randomised controlled trial of problem solving treatment, antidepressant medication, and combined treatment for major depression in primary care. *British Medical Journal*. 2000; 1(320):26–30. [PubMed: 10617523]
- Newman MG, Erickson T, Przeworski A, Dzus E. Self-help and minimal-contact therapies for anxiety disorders: Is human contact necessary for therapeutic efficacy? *Journal of Clinical Psychology*. 2003; 59(3):251–274. [PubMed: 12579544]
- Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: Is human contact necessary for therapeutic efficacy? *Clinical Psychology Review*. 2011; 31(1):89–103. [PubMed: 21130939]
- Nezu, AM.; D'Zurilla, TJ.; Zwick, ML.; Nezu, CM. Problem-solving therapy for adults.. In: Chang, EC.; D'Zurilla, TJ.; Sanna, LJ., editors. *Social problem solving: Theory, research, and training*. American Psychological Association; Washington, DC: 2004. p. 171-191.
- Ormrod JA, Kennedy L, Scott J, Cavanagh K. Computerised cognitive behavioural therapy in an adult mental health service: A pilot study of outcomes and alliance. *Cognitive Behaviour Therapy*. 2010; 39(3):188–192. <http://dx.doi.org/10.1080/16506071003675614>. [PubMed: 20485996]
- Oxman TE, Hegel MT, Hull JG, Dietrich AJ. Problem-solving treatment and coping styles in primary care for minor depression. *Journal of Consulting and Clinical Psychology*. 2008; 76(6):933–943. <http://dx.doi.org/10.1037/a0012617>. [PubMed: 19045962]
- Pittaway S, Cupitt C, Palmer D, Arowobusoye N, Milne R, Holttum S, Patrick H. Comparative, clinical feasibility study of three tools for delivery of cognitive behavioural therapy for mild to moderate depression and anxiety provided on self-help basis. *Mental Health in Family Medicine*. 2009; 6:145–154. [PubMed: 22477905]
- Proudfoot J, Goldberg D, Mann A, Everitt B, Marks I, Gray JA. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. *Psychological Medicine*. 2003; 33(2):217–227. <http://dx.doi.org/10.1017/S0033291702007225>. [PubMed: 12622301]
- Proudfoot J, Ryden C, Everitt B, Shapiro DA, Goldberg D, Mann A, Gray JA. Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: Randomised controlled trial. *British Journal of Psychiatry*. 2004; 185:46–54. [PubMed: 15231555]
- Reynolds DAI, Stiles WB, Grohol JM. An investigation of session impact and alliance in Internet based psychotherapy: Preliminary results. *Counselling and Psychotherapy Research*. 2006; 6:164–168. <http://dx.doi.org/10.1080/14733140600853617>.

- Richards D, Richardson T. Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review*. 2012; 32(4):329–342. <http://dx.doi.org/10.1016/j.cpr.2012.02.004>. [PubMed: 22466510]
- Richardson R, Richards DA, Barkham M. Self-help books for people with depression: The role of the therapeutic relationship. *Behavioural and Cognitive Psychotherapy*. 2010; 38:67–81. <http://dx.doi.org/10.1017/S1352465809990452>. [PubMed: 19995466]
- Spek V, Cuijpers PIM, Nyklíček I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: A meta-analysis. *Psychological medicine*. 2007; 37(3):319–328. [PubMed: 17112400]
- Steiner JL, Tebes JK, Sledge WH, Walker ML. A comparison of the structured clinical interview for DSM-III-R and clinical diagnoses. *Journal of Nervous and Mental Disease*. 1995; 183(6):365–369. [PubMed: 7798084]
- Stiles WB, Agnew-Davies R, Barkham M, Culverwell A, Goldfried MR, Halstead J, Shapiro DA. Convergent validity of the Agnew Relationship Measure and the Working Alliance Inventory. *Psychological Assessment*. 2002; 14(2):209–220. <http://dx.doi.org/10.1037/1040-3590.14.2.209>. [PubMed: 12056083]
- Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: A systematic review. *Journal of Medical Internet Research*. 2012; 14(4)
- Swift JK, Greenberg RP. Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*. 2012; 80(4):547–559. <http://dx.doi.org/10.1037/a0028226>. [PubMed: 22506792]
- Thornett AM, Mynors-Wallis LM. Credibility of problem-solving therapy and medication for the treatment of depression among primary care patients. *Medical Science Monitor*. 2002; 8(3):193–196.
- Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: Results from the national comorbidity survey replication. *Archives of General Psychiatry*. 2005; 62(6):629–640. <http://dx.doi.org/10.1001/archpsyc.62.6.629>. [PubMed: 15939840]
- Ware JE Jr. SF-36 health survey update. *Spine*. 2000; 25(24):3130–3139. [PubMed: 11124729]
- Ware, JE.; Kosinski, M. SF-36 physical & mental health summary scales: A manual for users of version 1. Quality Metric Inc.; Lincoln, RI: 2001.
- Williams JW Jr, Barrett J, Oxman T, Frank E, Katon W, Sullivan M, Cornell J, Sengupta A. Treatment of dysthymia and minor depression in primary care: A randomized controlled trial in older adults. *Journal of the American Medical Association*. 2000; 284:1519–1526. <http://dx.doi.org/10.1001/jama.284.12.1519>. [PubMed: 11000645]
- Williams JR Jr, Stellato CP, Cornell J, Barrett JE. The 13- and 20-item Hopkins Symptom Checklist Depression Scale: Psychometric properties in primary care patients with minor depression or dysthymia. *The International Journal of Psychiatry in Medicine*. 2004; 34(1):37–50.
- Williams JB, Kobak KA. Development and reliability of a structured interview guide for the Montgomery–Åsberg Depression Rating Scale (SIGMA). *The British Journal of Psychiatry*. 2008; 192(1):52–58. [PubMed: 18174510]
- Zanarini MC, Frankenburg FR. Attainment and maintenance of reliability of axis I and axis II disorders over the course of a longitudinal study. *Comprehensive Psychiatry*. 2001; 42(5):369–374. [PubMed: 11559863]
- Zanarini MC, Skodol AE, Bender D, Dolan R, Sanislow C, Schaefer E, Gunderson JG. The Collaborative Longitudinal Personality Disorders Study: Reliability of axis I and II diagnoses. *Journal of Personality Disorders*. 2000; 14(4):291–299. [PubMed: 11213787]
- Zimmerman M, Posternak MA, Chelminski I. Derivation of a definition of remission on the Montgomery-Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. *Journal of Psychiatric Research*. 2004; 38:577–582. [PubMed: 15458853]
- Zuroff DC, Blatt SJ, Sotsky SM, Krupnick JL, Martin DJ, Sanislow CA, Simmens S. Relation of therapeutic alliance and perfectionism to outcome in brief outpatient treatment of depression.

Journal of Consulting and Clinical Psychology. 2000; 68(1):114–124. <http://dx.doi.org/10.1037/0022-006X.68.1.114>. [PubMed: 10710846]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

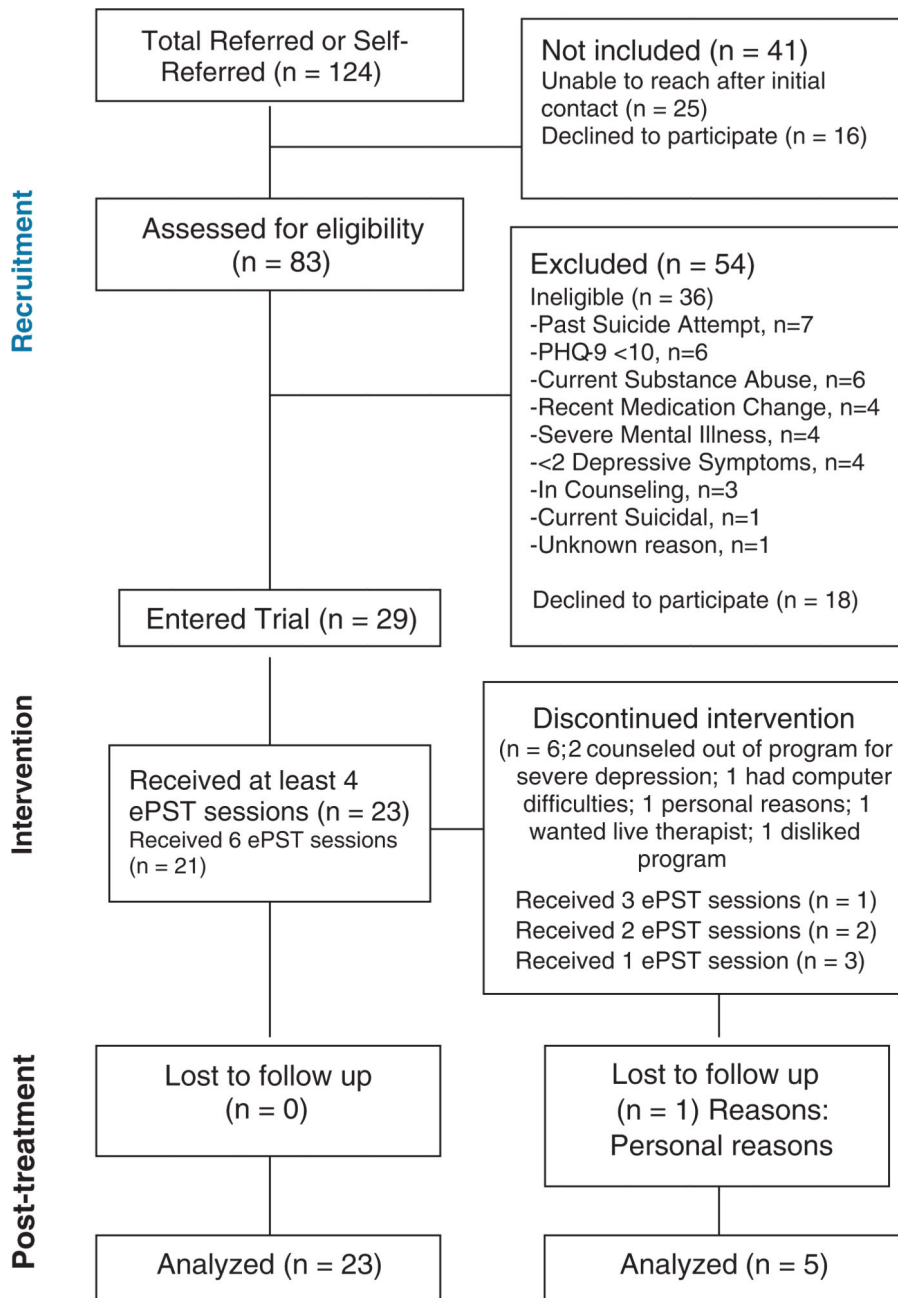


FIGURE 1.
ePST™ Participant Flow Through the Study.

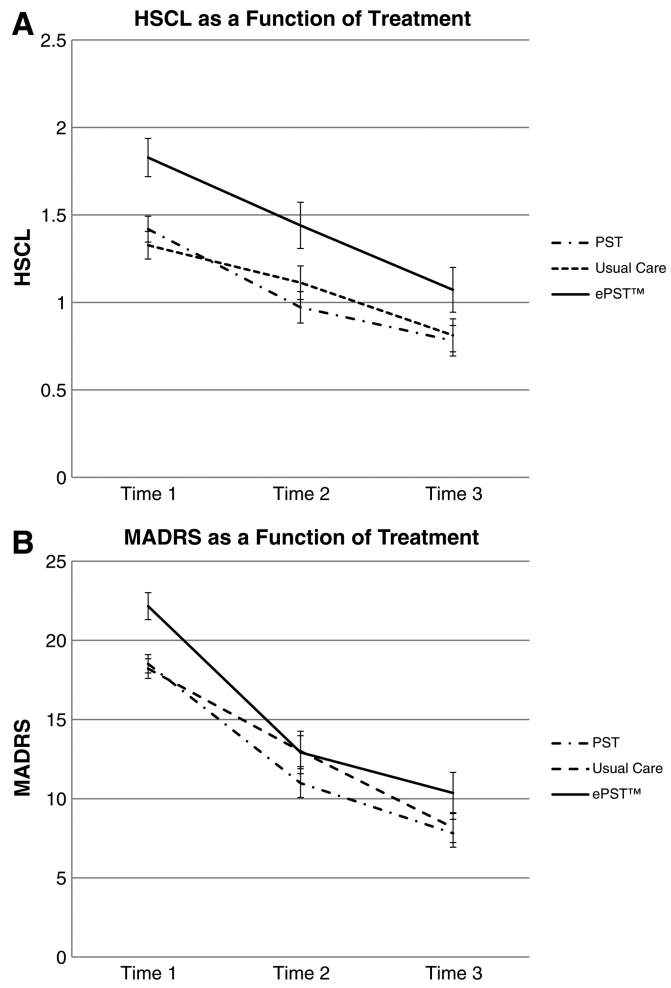


FIGURE 2.

A: HSCL as a Function of Treatment. B: MADRS as a Function of Treatment.

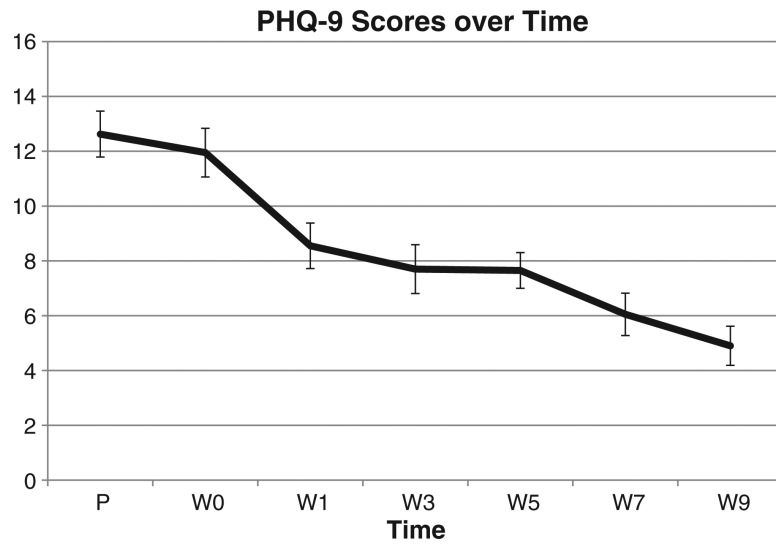


FIGURE 3.
PHQ-9 Scores over Time.

Table 1

Number of Treatment Sessions, Type of Therapy, Measures, and Assessment Time Points for *ePST*TM and Other Studies Used for Relevant Comparisons

Study	Number of Sessions/Timeframe	Therapy Type	Measures used	Assessment Timepoints used for Comparison
<i>ePST</i> TM	6 weekly/biweekly over 9 weeks	<i>ePST</i> TM	ARM HSCL, MADRS PHQ-9 HAMD	Week 4, 10 Week 0, 4, 10 Week -1, 0, 1, 3, 5, 7, 9 Week 0
Ormrod et al., 2010	9 weekly	Computer-based CBT, <i>BtB</i>	ARM	Week 1-9
Agnew-Davies et al., 1998	8 weekly	Live CBT	ARM	Week 1-8
Oxman et al., 2008	6 weekly/biweekly over 9 weeks	Live PST	HSCL HAMD	Week 0, 4, 10 Week 0
Mynors-Wallis et al., 1995	6 weekly/biweekly over 12 weeks	Live PST	BDI HAMD	Week 0, 6, 12 Week 0

Note. *ePST*TM = Electronic Problem Solving Treatment; ARM = Agnew Relationship Measure; HSCL = Hopkins Symptom Checklist-20 Item Depression Scale; MADRS = Montgomery-Asberg Depression Rating Scale; PHQ-9 = Patient Health Questionnaire-9 Item Depression Scale; HAMD = Hamilton Rating Scale for Depression-17 Item; BDI = Beck Depression Inventory.

Table 2

Means and Standard Deviations of All Measures for the Time Points They Were Delivered (Baseline, 4 Weeks, 10 Weeks)

Measure	Baseline	4 Weeks	10 Weeks
	Mean (SD)	Mean (SD)	Mean (SD)
Eligibility			
HAMD	16.8 (4.6)		
Feasibility/Process			
SUS		78.8 (13.4)	80.3 (17.0)
AST		5.3 (1.1)	5.5 (1.3)
CQ		6.7 (1.9)	7.5 (2.0)
ARM			
Bond Scale		6.0 (0.6)	6.2 (0.7)
Partnership Scale		5.3 (1.1)	5.5 (1.0)
Confidence Scale		5.6 (0.8)	5.9 (0.8)
Openness Scale		5.3 (1.0)	5.6 (0.8)
Clinical Outcomes			
HSCL	1.9 (.58)	1.4 (0.7)	1.1 (0.8)
MADRS	22.7 (4.7)	13.3 (6.6)	10.4 (6.6)
SF-36, Role Physical	77.8 (4.7)	62.0 (23.6)	66.7 (41.6)
SF-36, Role Emotional	34.5 (33.9)	28.0 (34.3)	56.8 (39.0)
SF-36, Social Functioning	48.3 (20.2)	60.5 (23.6)	69.0 (24.6)

Note. HAMD = 17-item Hamilton Depression Rating Scale, SUS = System Usability Scale, AST = Acceptability of Self-Guided Treatment Scale, CQ = Credibility Questionnaire, ARM = Agnew Relationship Measure, HSCL = Hopkins Symptom Checklist 20-Item Depression Scale, MADRS = Montgomery-Asberg Depression Rating Scale, SF-36 = Medical Outcomes Short Form – 36.

Table 3

Means and Standard Deviations for ARM Subscales (Bond, Partnership, Confidence, and Openness) for *ePST*TM, Live CBT for Depression, and *BtB*

ARM Subscale	<i>ePST</i> TM 4 weeks Mean (SD)	CBT Live 4 weeks (Agnew-Davies et al., 1998) Mean (SD)	<i>ePST</i> TM 10 weeks Mean (SD)	CBT Live 8 weeks (Agnew-Davies et al., 1998) Mean (SD)	<i>BtB</i> 9 weeks (Ormrod et al., 2010) Mean (SD)
Bond	5.99 (0.64)	5.85 (0.82)	6.13 (0.74)	5.95 (0.81)	5.16 (0.47)
Partnership	5.32 (1.11)	5.90 (0.83)	5.53 (1.01)	5.98 (0.79)	4.97 (0.54)
Confidence	5.59 (0.80)	5.80 (0.80)	5.92 (0.80)	5.88 (0.85)	4.95 (0.62)
Openness	5.30 (0.98)	5.30 (1.09)	5.63 (0.83)	5.55 (1.04)	4.60 (0.66)

Note. Due to unacceptably low inter-item correlations on the Client Initiative subscale these comparisons were not conducted. *BtB* = *Beating the Blues* (Ormrod et al., 2010).

Table 4Demographics and Baseline Depression Symptoms for *ePST*TM and Comparison Sample Participants

	<u>Usual Care (Oxman et al., 2008; n = 69)</u>	<u>Live PST (Oxman et al., 2008; n = 72)</u>	<u><i>ePST</i>TM (n = 29)</u>	<i>p</i>
	% or Mean (SD)	% or Mean (SD)	% or Mean (SD)	
Age	56 (15.5)	55 (16.5)	53 (12.6)	> .25
Female	58%	58%	79%	.10
Married	62%	65%	55%	> .25
Employed	70%	58%	66%	> .25
Income > \$40,000	68%	68%	52%	.24
Attended College	52%	56%	83%	.02
Baseline HAMD Score	13.4 _a (3.1)	14.1 _a (3)	16.8 _b (4.6)	< .001
Baseline HSCL Score	1.4 _a (.58)	1.5 _a (.62)	1.9 _b (.58)	.001
Baseline MADRS Score	18.6 _a (4.3)	18.7 _a (4.5)	22.7 _b (4.7)	< .001
Baseline SF36 Role Physical Scale	73.2 (24.1)	75.1 (23.6)	77.8 (22.2)	> .25
Baseline SF36 Role Emotional Scale	43.5 (38.5)	39.4 (39.7)	34.5 (33.9)	> .25
Baseline SF36 Social Functioning Scale Score	63.9 _a (20.7)	67 _a (23.7)	48.3 _b (20.2)	< .001

Note. The final column displays *p*-values from Pearson chi-square (for dichotomous variables) and one-way ANOVA (for continuous variables). For continuous variables with significant *F*-test, row entries that do not share a common subscript differ at *p* < .05 using Bonferroni corrected *t*-tests.