Behaviour Research and Therapy 72 (2015) 72–80

Contents lists available at ScienceDirect

Behaviour Research and Therapy

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

Comparing human and automated support for depression: Fractional factorial randomized controlled trial

Saskia M. Kelders*, Ernst T. Bohlmeijer, Wendy T.M. Pots, Julia E.W.C. van Gemert-Pijnen

University of Twente, Center for eHealth and Wellbeing Research, Department of Psychology, Health and Technology, Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 29 October 2014
Received in revised form
25 June 2015
Accepted 29 June 2015
Available online 6 July 2015

Keywords:
Web-based intervention
Depression
Support
Adherence
Persuasive technology

ABSTRACT

Web-based interventions for people with depressive symptoms are needed and show promising effects. However, it is a consistent finding that human support is needed and this makes implementation costly. This study investigates the adherence and effectiveness of a human-supported and automated-supported web-based intervention for people with mild to moderate depressive symptomatology, and studies the impact of four persuasive technology components. People with mild to moderate depressive symptoms according to the Center of Epidemiological Studies depression scale self-report questionnaire were included, but no diagnosis was made for the study. Participants ($n = 239$) were randomized into one of eight intervention arms, where each level of each component is present in half of the intervention arms. On clinical outcomes, there was a significant interaction effect between support condition and time, but there was no difference on the extent of improvement from baseline to follow-up, only a difference in the time-path of improvement. Effect sizes from baseline to follow-up were 0.89 for automated and 1.00 for human support. There was no significant difference on adherence between support condition. We conclude that an automated-supported web-based intervention for treatment of depression with persuasive technology may achieve similar adherence and effectiveness as the same intervention with human support.

1. Introduction

Web-based interventions for people with mild to moderate depressive symptomatology are needed and show promising effects (Andersson & Cuijpers, 2009; Andersson et al., 2005; Barak, Hen, Boniel-Nissim, & Shapira, 2008; Jorm & Griffiths, 2006; Kaltenthaler, Parry, Beverley, & Ferriter, 2008; Musiat & Tarrier, 2014; Spek et al., 2007). However, it is a consistent finding that human support is necessary to ensure adherence (i.e. following the intervention protocol) and to increase the effects (Andersson & Cuijpers, 2008, 2009; Hilvert-Bruce, Rossouw, Wong, Sunderland, & Andrews, 2012; Musiat & Tarrier, 2014; Spek et al., 2007). This counselor involvement makes large scale implementation still a costly business. Therefore, a major challenge is to develop web-based interventions that are as effective as human-supported interventions, but have less counselor involvement.

However, questions remain on what support is needed within web-based interventions. E.g. does the effectiveness of support stem from the actual involvement of a counselor, or (partly) from other factors, such as having clear deadlines within the treatment (Andersson & Cuijpers, 2009) or contact before and after treatment (Johansson & Andersson, 2012)? Moreover, research shows that guidance of a therapist is not essential to produce significant benefits as long as nonguidance contact is provided, e.g. from a technician (Talbot, 2012; Titov et al., 2010). This poses the question whether support needs to come from a human, or whether (some of) the support can be automated (Andersson & Cuijpers, 2009; Johansson & Andersson, 2012).

Studies have shown that automated support can be effective, although less than human support or with lower adherence rates (Furmark et al., 2009; Morgan, Jorm, & Mackinnon, 2012; Titov, Andrews, Choi, Schwencke, & Johnston, 2009). It seems that automated-supported interventions need to be enhanced to reach similar effectiveness and adherence as human-supported interventions. Persuasive technology may provide the means to enhance these interventions (Fogg, 2003; Oinas-Kukkonen & Harjumaa, 2009). Promising persuasive technology features are text-messages, interaction, tailoring and personalization. Review

* Corresponding author. University of Twente, Room Cu B114, PO Box 217, 7500AE Enschede, The Netherlands.
E-mail address: s.m.kelders@utwente.nl (S.M. Kelders).

http://dx.doi.org/10.1016/j.brat.2015.06.014
0005-7967/

studies have shown that interventions that include text messages are more effective than interventions that do not include text messages (Webb, Joseph, Yardley, & Michie, 2010) and that reminders increase the effect and adherence of web-based interventions (Fry & Neff, 2009). Furthermore, interactivity has been shown to increase adherence and effectiveness of web-based interventions (Hurling, Fairley, & Dias, 2006; Ritterband et al., 2006). Also tailoring (i.e. adapting content to a particular group of people (de Vries & Brug, 1999)) has been shown to be effective in health behavior change interventions (Noar, Benac, & Harris, 2007; Strecher et al., 2008). Lastly, personalization is proposed to increase the persuasiveness of technology (Fogg, 2003; Oinas-Kukkonen & Harjumaa, 2009). This can be achieved by adaptation (e.g. a shopping website shows recommended products based on the history of the user) and adaptability (e.g. a website provides the opportunity to show a self-chosen picture on your personal page) (Knutov, De Bra, & Pechenizkiy, 2009; Velsen, 2011).

For the present study, the web-based intervention 'Living to the full' was developed based on an evidence-based intervention for adults with mild to moderate depressive symptomatology (Fledderus, Bohlmeijer, Pieterse, & Schreurs, 2011; Kelders, Pots, Oskam, Bohlmeijer, & van Gemert-Pijnen, 2013). The intervention is based on Acceptance and Commitment Therapy (ACT) and targets psychological flexibility; the ability to accept thoughts and feelings in order to create space for evaluating and pursuing valued life activities (Hayes, Strosahl, & Wilson, 1999). ACT has been recognized by the APA as an evidence-based treatment of depression. The intervention is aimed at adults with mild to moderate depressive symptomatology, because having clinically relevant depressive symptoms is the most important risk factor for developing a depression (Cuijpers & Smit, 2004) and web-based interventions can be a good strategy to reach this target group (Andersson & Cuijpers, 2008; Cuijpers, van Straten, Warmerdam, & van Rooy, 2010). The web-based intervention includes persuasive technologies that have been found to impact adherence and effectiveness of interventions. Furthermore, a human-supported and an automated-supported version were created. Additionally, process measures were included to assess how participants evaluate the intervention.

2. Method

2.1. Experimental design

In order to compare the human-supported intervention with the automated-supported intervention and to study the relative impact of each persuasive technology component, the screening phase of the Multiphase Optimization Strategy (MOST) was used (Collins, Murphy, & Strecher, 2007). The purpose of this methodology is efficiently selecting active components of interventions. Most behavioral interventions can be seen as consisting of multiple components. Some of these components are part of the intervention itself (e.g. the program content consisting of text and exercises) and other components are related to the delivery of the program (e.g. feedback by an expert or through an automated system). Standard randomized controlled trials can only investigate the effect of the intervention as a whole, i.e. as the sum of the components. However, it may be that some components are very effective and others provide little value or even diminish the effect of other components. It is important to identify and select the active components (the components that provide the wanted effects) and this is the purpose of the MOST screening phase. This is done through randomized experimentation by a fractional factorial trial. Numerous authors have suggested using this methodology to better assess active components in web-based interventions (e.g. Glasgow, 2007; Norman, 2008; Morrison, Yardley, Powell, & Michie, 2012).

Following the MOST screening phase, we have identified a number of components of which we would like to know whether they are active components. These components are: support (human or automated feedback); text message coaching (present or absent); interaction (high or low); tailoring of success stories (high or low); and personalization (high or low). A full factorial design to investigate these five components with two levels each would require a study with $2^5 = 32$ arms (component combinations) which would not be feasible for this study. However, according to MOST, a fractional factorial design provides an alternative: to be able to answer the research question, not all the arms are needed (Collins, Murphy, Nair, & Strecher, 2005). For this study, the main effects of the components are most important. Furthermore, there is no theory or hypothesis to support an effect of 2-way interactions (the effect of one component influencing the effect of another component), therefore we assume that the impact of these interactions is negligible. Based on these assumptions, we identified a balanced 8 arm fractional factorial design which allowed us to compare the human-supported web-based intervention with the automated-supported web-based intervention and to screen for the effects of the other four components (a 'resolution III' design (Box, Hunter, & Hunter, 2005), see Appendix A). Each arm of the trial includes a different combination of the levels of the components. E.g. participants randomized in the first arm receive the intervention with automated support, text messages, high interaction, high tailoring and high personalization. Participants in the second arm receive the intervention with automated support, text messages, low interaction, low tailoring and low personalization. The design is balanced, which means that each level of each component is present in half of the intervention arms (e.g. four intervention arms include automated support and the other four arms include human support). For the analysis of the effects of the levels of each component, we compared all participants that received a certain level of a component with the participants that received the other level of the component (e.g. comparing all participants who received human feedback with all participants who received automated feedback).

2.2. Recruitment and participants

Inclusion criteria were an age of 18 year or older and mild to moderate depressive symptoms (>9 and <39 on the Center of Epidemiological Studies – depression scale; CES-D) (Radloff, 1977). Exclusion criteria were severe depressive symptomatology and/or severe anxiety symptoms (more than 1 standard deviation above the population mean on the CES-D (cut-off score 39) (Bouma, Ranchor, Sanderman, & Van Sonderen, 1995)) and/or on the Hospital Anxiety and Depression Scale – anxiety subscale (HADS-A (Zigmond & Snaith, 1983); cut-off score 15 (Olsson, Mykletun, & Dahl, 2005)), because for these people more comprehensive diagnosis and psychiatric treatment is warranted. Other exclusion criteria were: receiving psychological or psycho-pharmacological treatment within the last 3 months; having less than 3 h per week time to spend on the web-based intervention; poor Dutch language skills.

2.3. Procedure

Participants were recruited through advertisements in Dutch newspapers between February and March 2011. Interested people visited the study website. After reading information about the study, informed consent was obtained from the participant through a checkbox and a pop-up screen to check whether they were sure to

give informed consent. Participants then filled out an online screening questionnaire and were instantly informed whether they fulfilled the inclusion criteria. People who fulfilled the inclusion criteria were emailed a link to the online baseline questionnaire. A total of 239 respondents fulfilled the inclusion criteria, completed the online baseline questionnaire and were automatically randomized to one of eight intervention arms. All participants could start the web-based intervention on the same day (25 March). It was not possible to blind respondents to their randomized arm. However, they had no in-depth knowledge of the other arms. Participants received an email with an invitation to the online post intervention questionnaire three months after the start of the intervention period. Six months after the start of the intervention period, participants received an emailed link to the online follow-up questionnaire. Up to two automated email reminders were sent to the participants when not filling out a questionnaire.

2.4. Intervention

The web-based intervention 'Living to the full' included 9 chronological lessons and is based on Acceptance and Commitment Therapy. Each module included text, online and offline exercises, and metaphors to illustrate the six processes of ACT: acceptance (active and nonjudgmental embrace of experiences), cognitive defusion (letting go of entanglement with negative and unwanted thoughts), contact with the present moment, self as context (experiencing thoughts and feelings from a nonjudgmental and reflexive point of view), evaluating values in different life domains, and commitment to actions based on these values (Hayes et al., 1999). Participants were instructed to complete 1 lesson per week, but had 12 weeks in total to complete the 9 lessons. A detailed description of the intervention can be found in Appendix B. More information on the content of the intervention can be found in (Pots et al., in press).

2.5. Components

Following is a short description of the implementation of the five components in the intervention. More detailed information can be found in Appendix B.

2.5.1. Support

The source of support was either human or automated. To isolate the effect of the source of support, both conditions were designed to be comparable regarding length of feedback messages, tailored content, and presentation (including a picture of the [virtual] counselor). On the other hand, both sources of support have different possibilities. Human support provides an increased possibility for substantive interaction which is difficult to achieve with automated support. Conversely, automated support provides an increased possibility for timely or instant feedback, which is difficult with human support. To maintain these innate differences, participants in the human support condition had the opportunity to ask questions to their counselor, and participants in the automated support condition received one additional instant feedback message per lesson. Feedback messages were provided within the application and participants received automated email messages when feedback was received.

2.5.2. Text message coaching

The text messages in the arms that included text message coaching were written before the study started by the researchers and the content was based on the results of the development study of the intervention (Kelders et al., 2013). Each week, three text messages were sent containing motivational, mindfulness and content-related information.

2.5.3. Interaction

The high interaction arms contained additional multimedia and interactive material in the form of short movies, interactive exercises and multimedia presentations of metaphors. The low interaction condition did not include these elements.

2.5.4. Tailoring of success stories

The intervention contained a success story for each of the lessons of the intervention. For the high tailored arms, each success story was tailored on four of the aspects: gender, age, marital status, daily activity, most prominent symptom, reason for participating in the web-based intervention. E.g. in lesson one participants in the high tailored arms were shown a success story of a person of the same gender and age group, who has the same symptoms as they have and the same reason for participating in the web-based intervention. The stories were tailored to a different combination of aspects each week and not on all aspects to maintain the credibility of the stories. In the low tailored arms, a standard success story was presented each week.

2.5.5. Personalization

In this study, the high personalization arms included personalized content that is adapted (the system shows the motto and picture selected by the participant; the system shows the most important values selected by the participant) and adaptable (possibility to create a personal 'top 5' aspects from the course that the participant found most important).

2.6. Measures

Adherence to the web-based intervention was measured objectively by log files. Participants reaching lesson 9 were classified as adherent. Depressive symptoms were measured with the CES-D (20 items, score 0–60; higher scores mean more depressive symptoms (Haringsma, Engels, Beekman, & Spinhoven, 2004; Radloff, 1977) at baseline, post intervention and follow-up. Anxiety symptoms were measured with the HADS-A (7 items, score 0–21; higher scores mean more anxiety symptoms (Zigmond & Snaith, 1983; Spinhoven et al., 1997)) at baseline, post intervention and follow-up. The process measures task enjoyment, involvement, trust and satisfaction with the web-based intervention were measured at post intervention. Task enjoyment is an important component of social cognitive theories of achievement and intrinsic motivation and is seen as a mediator between achievement motivation and performance, and was measured by 5 items (Tauer & Harackiewicz, 1999). Involvement is defined as 'a person's perceived relevance of the object based on inherent needs, values and interests' and may be related to adherence. Involvement was measured with the short version of the Personal Involvement Inventory (10 items) (Zaichkowsky, 1994). Trust is widely considered to be important in whether people decide to accept information and advice on their health. Specifically, trust in the organization and in technology are seen as important factors that determine whether a person will use a website (Velsen, 2011). Trust was measured with two constructs: trust in the organization (4 items) and trust in the technology (4 items) (Velsen, 2011). Satisfaction with a web-based intervention might predict adherence and was measured with 4 items on user friendliness, usefulness, recommending to others, and willingness to continue using the web-based intervention (Kelders, Van Gemert-Pijnen, Werkman, Nijland, & Seydel, 2011). For all process measure a mean score was calculated (range 1–7; for satisfaction range 1–5), where a higher score is more favorable towards the intervention.

2.7. Data analysis

Statistical analyses were done using SPSS 18 (IBM, USA). All tests were two-tailed. Differences between randomized conditions and between responders and non-responders were investigated using one-way analyses of variance (ANOVA) and χ^2 tests. Missing data on clinical measures (CES-D and HADS-A) were imputed with the expectation-maximization (EM) method in SPSS 18. This method estimates the unmeasured data based on maximum likelihood estimates using observed data in an iterative process (Dempster, Laird, & Rubin, 1977). For this estimation, observed data on CES-D, HADS-A, gender, age, education, lesson reached and support condition were used. To examine differences on clinical outcome measures between the different levels of the components, Repeated Measures ANOVA with intention-to-treat data were used. All components were used as covariates. Cohen's d based on pooled standard deviations was used to calculate within subjects effect sizes (effect size of 0.56–1.2 was considered large, 0.33–0.55 as moderate, and less than 0.33 as small; (Lipsey & Wilson, 1993)). Cohen's d for within-subject effect sizes was corrected for dependence among means by using the correlation between the two means (Morris & DeShon, 2002). The sample size was considered satisfactory, since power calculations showed that there was a chance of over 80% to detect a difference of an effect size of 0.37 with an alpha of 0.05. To examine differences on adherence, per-protocol regression analyses were used with all components added to the model (i.e. only participants that have used the intervention at least once have been included in the analyses). To investigate the dose – response relationship (i.e. examine whether more usage of the intervention leads to more favorable outcomes), regression analyses were used with the clinical outcomes (CES-D and HADS-A on post intervention and follow-up) as dependent variables and adherence and lesson reached as predictor variables. To examine differences on process outcomes, ANOVA and regression analyses were used with per-protocol data of completers (i.e. participants that have used the intervention and filled out the post intervention questionnaire).

3. Results

3.1. Randomization

Due to a programming error in the randomization procedure, the number of participants in each arm differed from 11 to 53 (see Appendix A). To cope with this difference, we analyzed the data using all components as covariates (for ANOVA) or predictors (for regression analyses). The number of participants randomized in the human-supported and automated-supported group is fairly equal (113 and 126 respectively, see Fig. 1) and only the tailoring component shows a large difference in the number of participants between levels (88 in the low tailored condition and 151 in the high tailored condition). The error in randomization slightly influenced the power: for tailoring (least powered due to the error) there was a chance of 80% to detect a difference of an effect size of 0.38 with an alpha of 0.05.

3.2. Response rates

Of the 239 participants, 137 participants completed the post-intervention questionnaire and 135 participants completed the follow-up questionnaire (Fig. 1). There were no differences in response rates between the levels of the components. Males, lower educated participants and younger participants were more often drop-outs on post-intervention ($\chi^2_1 = 5.452$, $p = .02$; $\chi^2_2 = 13.703$, $p = .001$; $F_{1,\ 237} = 3.905$, $p = .049$, respectively). Moreover, participants that did not adhere to the intervention were more often drop-outs on post-intervention and follow-up ($\chi^2_1 = 90.458$, $p < .001$; $\chi^2_1 = 94.990$, $p < .001$, respectively).

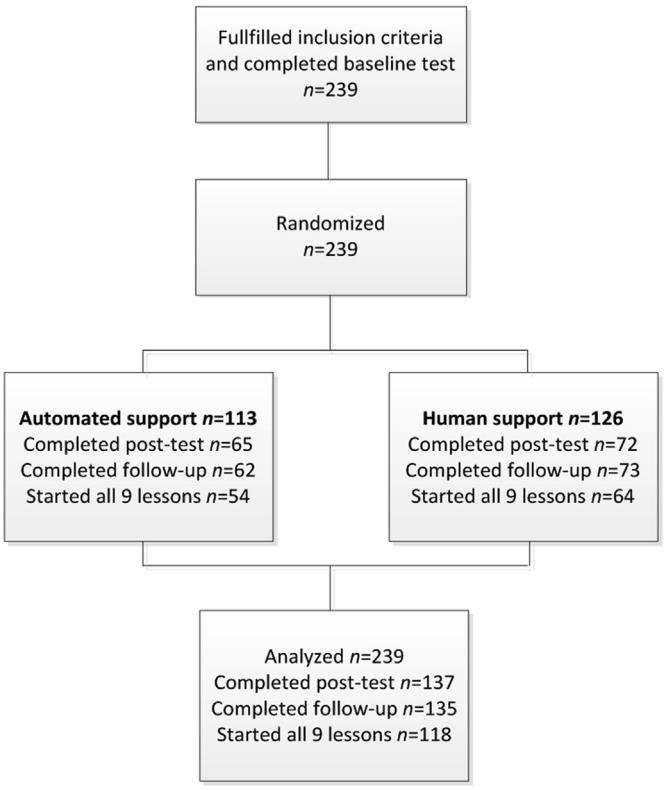

Fig. 1. Consort diagram.

3.3. Participant characteristics

Baseline demographics of participants by support condition are presented in Table 1. Possible differences in baseline characteristics were examined by the five experimental conditions. Of the 5 × 8 comparisons, there were three significant differences at the $p < 0.05$, where females more often received text message coaching, more often received high interaction and less often received high tailored success stories.

3.4. Adherence

The 239 participants completed on average 5.9 lessons within the intervention period (mode = 8, s.d. = 3.6). Of the participants, 33 (14%) did not start the intervention, and 118 (49%) completed all 9 lessons and therefore adhered to the intervention. A logistic regression showed that none of the components of the intervention significantly predicted adherence (all p-values >0.10).

3.5. Clinical outcomes

Data on the effectiveness of the intervention is presented in Table 2. A Repeated Measures ANOVA on the CES-D and HADS-A measures on baseline, post intervention and follow-up by intervention components, showed a significant effect of time and a significant interaction effect of time × support (Table 2). None of the other interactions were significant. For both outcome measures there was a significant time effect, showing that all groups significantly improved. Within subjects effect sizes for baseline to post-

Table 1
Participant characteristics by support condition.

	Automated support (n = 126)	Human support (n = 113)	Total (n = 239)
Age Mean (s.d.)	44.1 (11.8)	45.5 (12.8)	44.9 (12.3)
Gender percent (number)			
Male	31.0 (35)	27.8 (35)	29.3 (70)
Female	69.0 (78)	72.2 (91)	70.7 (169)
Ethnicity percent (number)			
Dutch	88.5 (100)	93.7 (118)	91.2 (218)
Other	11.5 (13)	6.3 (8)	8.8 (21)
Education level percent (number)			
High	69.0 (78)	63.5 (80)	66.1 (158)
Middle	22.1 (25)	30.2 (38)	26.4 (63)
Low	8.8 (10)	6.3 (8)	7.5 (18)
Marital status percent (number)			
Married	39.8 (45)	32.5 (41)	36.0 (86)
Divorced	19.5 (22)	23.0 (29)	21.3 (51)
Widowed	1.8 (2)	1.6 (2)	1.7 (4)
Unmarried	38.9 (44)	42.9 (54)	41.0 (98)
Daily activities percent (number)			
Paid job	62.8 (71)	63.5 (80)	63.2 (151)
Student	7.1 (8)	8.7 (11)	7.9 (19)
No job	30.1 (34)	27.8 (35)	28.9 (69)
CES-D Mean (s.d.)	24.3 (7.1)	25.6 (6.8)	25.0 (7.0)
HADS-A Mean (s.d.)	9.6 (2.6)	9.8 (2.6)	9.7 (2.6)

CES-D, Center of Epidemiological Studies – Depression scale; HADS-A, Hospital Anxiety and Depression Scale – Anxiety subscale.

intervention are moderate for automated support and large for human support and for both groups together. For baseline to follow-up, all effect sizes are large. The within subject effect sizes for the levels of the other components are similar, with small differences between the levels (Appendix C). Furthermore, on both outcome measures, there was a significant time × support interaction effect, although within-subjects contrasts show that this is not a linear effect, but a quadratic effect (CES-D: $F_1 = 12.370$; $p = .001$; HADS-A: $F_1 = 14.790$; $p < .001$; Fig. 2). One-way ANOVAs revealed a significant difference between the support conditions on post intervention on HADS-A ($F_{1,\ 237} = 4.716$, $p = .031$), but on follow-up there is no significant difference between support conditions. Looking at the effect sizes of the between groups differences, we see that on post-intervention the difference is small in favor of the human support condition (CES-D: $d = 0.18$; HADS-A: $d = 0.28$), whether on follow-up the difference is small, but in favor of the automated support condition (CES-D: $d = -0.13$; HADS-A: $d = -0.11$). Repeated Measures ANOVAs for both groups separately, showed that only the automated support condition significantly improved between post intervention and follow up (CES-D: $F_{1,\ 112} = 19.841$; $p < .001$; HADS-A: $F_{1,\ 48} = 7.590$; $p < .01$). All reported analyses are intention to treat, but per protocol analyses showed the same results.

3.6. Dose – response relationship

Adherence and lesson-reached significantly predicted improvement on clinical outcomes (CES-D and HADS-A on post intervention and follow-up). All regression analyses were significant with $p < .001$ and β ranging from 0.242 to 0.422, which supports the dose–response relationship. To illustrate this finding: the effect sizes from baseline to follow-up for participants who completed up to five lessons are 0.64 and 0.33 for depressive and anxiety symptoms, respectively; for participants who completed all nine lessons these effect sizes are considerably larger: 1.20 and 1.12.

3.7. Process measures

Table 3 shows the scores of responders at post-intervention on all process measures. We investigated whether there were differences between adherers and non-adherers and between participants in the different levels of the components. In general, the intervention was evaluated positively. There were significant differences between adherers and non-adherers on task enjoyment, involvement and satisfaction ($F_{1,\ 134} = 17.644$, $p < .001$; $F_{1,\ 134} = 12.734$, $p = .001$; $F_{1,\ 133} = 7.694$, $p = .006$, respectively) and on involvement by support and by text message coaching ($F_{1,\ 132} = 4.411$, $p = .038$; $F_{1,\ 132} = 4.415$, $p = .038$, respectively).

4. Discussion

4.1. Outcomes independent of support condition

Overall, participants who received the intervention showed a reduction in depressive and anxiety symptoms. The effect sizes were moderate to large from baseline to post intervention and large

Table 2
Outcomes and repeated-measures ANOVA by support condition; intention to treat analyses (n = 239).

	Group	Score, mean (s.d.)			Anova: F		Effect size: d	
		Pre	Post	Follow-up	time	time × group	pre – post	pre – follow-up
CES-D	Auto (n = 126)	24.33 (7.11)	20.38 (7.98)	17.58 (8.10)	13.667***	4.150*	0.46	0.73
	Human (n = 113)	25.62 (6.81)	18.99 (7.32)	18.54 (7.32)			0.82	0.83
	Total (n = 239)	25.01 (6.97)	19.65 (7.65)	18.08 (8.22)			0.64	0.75
HADS-A	Auto (n = 126)	9.56 (2.58)	8.30 (2.95)	7.30 (2.97)	15.642***	7.638**	0.41	0.84
	Human (n = 113)	9.81 (2.57)	7.46 (3.01)	7.61 (2.96)			0.75	0.82
	Total (n = 239)	9.69 (2.57)	7.85 (3.00)	7.46 (2.96)			0.59	0.83

Scores are presented as mean (sd); *p < .05, **p < .01, ***p < .001; CES-D, Center of Epidemiological Studies – Depression scale; HADS-A, Hospital Anxiety and Depression Scale – Anxiety subscale.

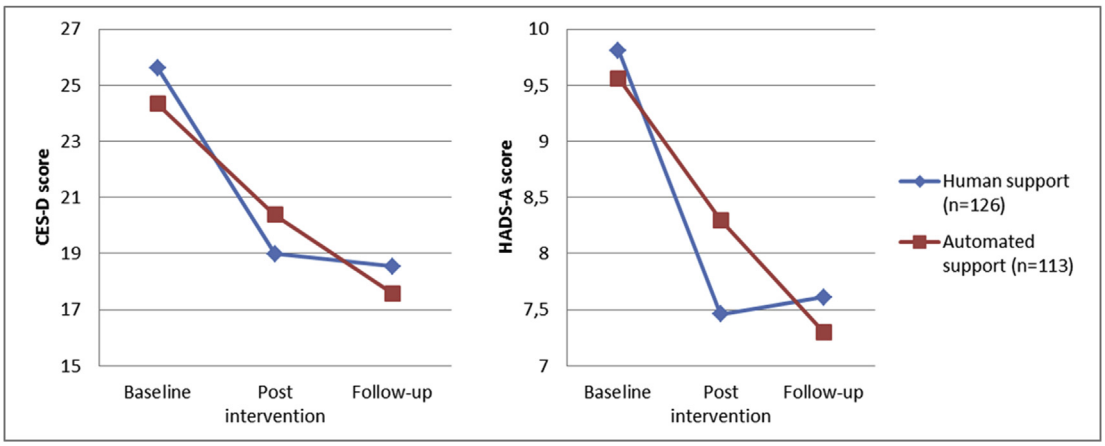

Fig. 2. Time × support interaction effect on CES-D and HADS-A.

Table 3
Scores of responders on process measures and differences between adherence-groups and levels of components, per protocol analyses.

	Task enjoyment mean (s.d.)	Involvement M mean (s.d.)	Trust in the organization mean (s.d.)	Trust in technology mean (s.d.)	Satisfaction mean (s.d.)
All participants	5.83 (1.16)	5.69 (1.14)	5.94 (1.01)	5.38 (1.18)	4.21 (0.81)
Adherers (n = 104)	6.05 (0.94)***	5.88 (0.96)**	6.02 (0.95)	5.42 (1.21)	4.31 (0.69)**
Non-adherers (n = 31; n = 32)[a]	5.12 (1.50)	5.10 (1.45)	5.65 (1.15)	5.24 (1.11)	3.86 (1.07)
Component: Support					
Automated (n = 62)	5.70 (1.25)	5.50 (1.16)*	5.79 (1.03)	5.29 (1.08)	4.20 (0.83)
Human (n = 72; n = 71)[a]	5.99 (0.95)	5.90 (1.00)	6.08 (0.95)	5.46 (1.25)	4.24 (0.76)
Component: Text messages					
No (n = 64; n = 63)[a]	5.69 (1.16)	5.51 (1.12)*	5.82 (1.02)	5.35 (1.23)	4.18 (0.81)
Yes (n = 70)	6.00 (1.03)	5.90 (1.04)	6.05 (0.96)	5.41 (1.13)	4.25 (0.78)
Component: Interaction					
Low (n = 54)	5.67 (1.36)	5.51 (1.27)	5.83 (1.10)	5.25 (1.28)	4.14 (0.92)
High (n = 80; n = 79)[a]	5.97 (0.88)	5.85 (0.94)	6.02 (0.94)	5.47 (1.09)	4.28 (0.69)
Component: Tailoring					
Low (n = 82)	5.97 (1.03)	5.79 (1.04)	5.91 (0.93)	5.28 (1.21)	4.28 (0.77)
High (n = 52; n = 51)[a]	5.68 (1.20)	5.59 (1.17)	6.00 (1.10)	5.55 (1.10)	4.12 (0.83)
Component: Personalisation					
Low (n = 72; n = 71)[a]	5.80 (1.07)	5.65 (1.09)	6.08 (0.96)	5.40 (1.05)	4.26 (0.79)
High (n = 62)	5.92 (1.14)	5.78 (1.10)	5.79 (1.01)	5.37 (1.31)	4.17 (0.80)

[a] Due to missing data, the results on Trust in the organization, Trust in technology and satisfaction are based on the responses of in total 133 responders versus the responses of 134 responders on task enjoyment and involvement; *p < .05; **p < .01; ***p < .001.

from baseline to follow-up. This is comparable to the results of a study which compared this intervention (including human support and the persuasive technology components) to a minimal intervention and a waitlist control group (Pots et al., in press) and with the effectiveness of internet-delivered behavioral activation combined with ACT for depression (Carlbring et al., 2013). Furthermore, meta-analyses of the effectiveness of guided web-based interventions for depression show similar effect sizes (Andersson & Cuijpers, 2009; Spek et al., 2007). Despite our rigorous definition of adherence (finishing all 9 lessons), approximately half of the participants adhered to the intervention, which is similar to the average adherence of web-based interventions (Kelders, Kok, Ossebaard, & Van Gemert-Pijnen, 2012). Lastly, our study confirmed the dose–response relationship; adherence was significantly related to better clinical outcomes (see e.g. Donkin et al., 2011; Hedman et al., 2013).

4.2. Human and automated support

On clinical outcomes, there was a significant interaction effect between support condition and time, but this effect was quadratic, not linear. There was no difference on the extent of improvement from baseline to follow-up between participants who received automated support and participants who received human support. The quadratic interaction effect shows that there was a difference on the time-path of the improvement. For participants who received human support, the improvement almost exclusively took place during the intervention period; between post intervention and follow-up, the scores on outcome measures remained stable. Participants who received automated support showed less improvement during the intervention period, but the improvement carried on between post intervention and follow-up. A possible explanation for this difference can be found in agency: a sense of agency (the attribution of improvement to oneself instead of others, e.g. a therapist) is positively associated with the effectiveness of therapy (Adler, Skalina, & McAdams, 2008). Participants who received automated support had no therapist to attribute the improvement to and may therefore have attributed the improvement more to themselves, which may enhance the effectiveness of this condition. Another explanation may be that participants who received human support might not improve after the intervention period due to the sudden loss of support from the therapist. Participants who received automated support however, have no therapeutic alliance that can be lost and might therefore continue

to improve. However, these results remain difficult to interpret, because we cannot control for whether participants have received other treatments after the intervention period.

On adherence, there were no significant differences between automated and human support. This is contrary to earlier studies that found that the inclusion of human support increased adherence (e.g. Kelders et al., 2012). A possible explanation for the lack of difference in adherence, and additionally for the effectiveness of the automated-supported intervention, can be found in the implementation of persuasive technology in the current intervention. First, social presence was enhanced by using a photo of the automated counselor. This use of an avatar has been shown to have positive effects (Baylor, 2009). Second, the weekly feedback messages of the automated support were implemented to closely resemble a human feedback message: the message was tailored to answers given by the participant on the exercises of that week and the set-up, length and presentation of the message was similar to that of a feedback message of a human counselor. Third, the automated support included an additional instant feedback message per lesson to make use of the increased possibility for timely feedback in automated support.

4.3. Relative impact of components

Contrary to our hypotheses, all other components besides support, showed no differences on clinical outcomes or adherence. An explanation might be that the variations between the levels of the components were too small to show an effect. For example, the levels of the component tailoring, differed only within the success stories. Moreover, our study on usage and use patterns of this intervention (Kelders, Bohlmeijer, & van Gemert-Pijnen, 2013), showed that not all components (e.g. these success stories) were used frequently. This might have lessened the effect. Furthermore, research on the influence of different combinations of persuasive features is still in its infancy (Oinas-Kukkonen, 2010). Although there are studies that show that each feature on its own can increase adherence or effectiveness, it may be that these effects do not accumulate, i.e. the addition of another enhanced version of a different component does not further increase this effect. In this study, each study arm includes one or more enhanced components, which would explain the lack of difference between the variations.

4.4. Process measures

Overall, participants evaluated the intervention positively. Adherers evaluated the intervention more positively than non-adherers with significant differences on task enjoyment, involvement and satisfaction. Furthermore, on involvement there were significant differences between components, where participants who received human support, text message coaching and high interaction scored higher. It seems that involvement discriminates between how different interventions are received and might provide a linking pin between the intervention and adherence or effect (Kelders, 2015).

4.5. Limitations

Some limitations apply to this study. First, we did not perform a non-inferiority or equivalence trial. Therefore, we cannot state with certainty that automated and human support are equivalent on effectiveness and adherence. However, the differences between adherence (3% in favor of human support) and within-group effect sizes from baseline to follow-up (0.02 on anxiety symptoms in favor of automated support and 0.10 on depressive symptoms in favor human support) are so small that a study would need about 1300 respondents in each condition for the difference on depressive symptoms to reach statistical significance with a power of 0.80. Furthermore, the question is whether this difference is clinically relevant. Additionally, the study may not have been powered enough to detect differences between the components of persuasive technology. As this design is quite innovative, it is difficult to establish expected effect sizes for the different components and it may not be realistic to expect an effect size of 0.37. Moreover, this study showed that the differences in effect sizes between the levels of the components are very small (0.01–0.19) and the sample size would have to be very large for this difference to reach statistical significance.

Second, our findings may be less generalizable to people with more severe depressive complaints and to other clinical populations because of the exclusion of people who scored high on measures of depression or anxiety. However, with the inclusion of automated feedback, we feel that the participants represent the target group when implemented in regular care. Furthermore, we only used a self-report measure to assess depressive symptoms and we did not diagnose participants. Therefore, the findings may be less generalizable. Lastly, due to the programming error in the randomization procedure, there was a slight imbalance on gender in the conditions text-message coaching, interaction and tailoring of success stories. However, we have no indication that these features differ in effectiveness between males and females.

A third limitation is the substantial drop-out. Although a drop-out of 43% is not unusual in research into web-based interventions and drop-out is hard to prevent (Blankers, Koeter, & Schippers, 2010; Christensen, Griffiths, & Farrer, 2009), it can pose a serious problem. Drop-out in this study was larger than in other trials on guided internet interventions (e.g. Johansson et al., 2012), which may be explained by the full online nature of this study. Participants received automated reminders for assessments, but they had no face-to-face, phone or email contact with research staff during recruitment and assessments. It is likely that this influenced the drop-out in this study, but we feel this strategy was necessary because research suggests that these strategies also influence adherence (Kelders et al., 2012). Because adherence is one of the main outcomes in this study, we refrained from using strategies that might influence adherence and therefore the results of this study. We have dealt with the high drop-out rate by employing the EM method to impute missing data. This method has been shown to be one of the methods that leads to the most valid and reliable results (Blankers et al., 2010).

4.6. Implications

To the best of our knowledge, this is the first study in which a web-based intervention without human support, but enriched with evidence-based persuasive technologies is compared with the same intervention with human support. This study suggests that a web-based intervention with automated support can be as effective at six month follow-up and reach the same adherence as a web-based intervention with human support, though a second and larger study is needed to draw more robust conclusions. Nonetheless, this study shows the potential of automated support which may dramatically reduce the costs of treatment for people with mild to moderate depressive symptomatology and with higher education. This may make these web-based interventions easier to implement in regular care and may reach many more people with mild to moderate depressive symptomatology. This is an important goal as depression is becoming one of the diseases with the largest burden of disease. As a final note, this study does not implicate that human support is not needed nor useful within web-based interventions. It merely proposes to think about the necessity of

human support, and whether the valuable time of clinicians should be spend on supporting all people or only the people who need it the most. However, more adequately powered research on moderators and predictors of outcomes is needed to identify the people who need human support the most.

Financial support

The study was funded by the University of Twente and by the National Institute for Public Health and the Environment (RIVM, SOR project number S/270136). The study was approved by an independent medical ethics committee (METIGG; NL33619.097.10) and recorded in the Dutch primary trial register for clinical trials (NTR3007).

Conflicts of interest

All authors have been involved in the development of the web-based intervention.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.brat.2015.06.014.

References

Adler, J. M., Skalina, L. M., & McAdams, D. P. (2008). The narrative reconstruction of psychotherapy and psychological health. *Psychotherapy Research, 18*, 719–734.
Andersson, G., Bergstrom, J., Hollandare, F., Carlbring, P., Kaldo, V., & Ekselius, L. (2005). Internet-based self-help for depression: randomised controlled trial. *British Journal of Psychiatry, 187*, 456–461.
Andersson, G., & Cuijpers, P. (2008). Pros and cons of online cognitive-behavioural therapy. *British Journal of Psychiatry, 193*, 270–271.
Andersson, G., & Cuijpers, P. (2009). Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cognitive Behaviour Therapy, 38*, 196–205.
Barak, A., Hen, L., Boniel-Nissim, M., & Shapira, N. A. (2008). A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *Journal of Technology in Human Services, 26*, 109–160.
Baylor, A. L. (2009). Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364*, 3559–3565.
Blankers, M., Koeter, M. W. J., & Schippers, G. M. (2010). Missing data approaches in eHealth research: simulation study and a tutorial for nonmathematically inclined researchers. *Journal of Medical Internet Research, 12*.
Bouma, J., Ranchor, A. V., Sanderman, R., & Van Sonderen, E. (1995). *Het meten van symptomen van depressie met de CES-D: Een handleiding* [*Measuring Symptoms of Depression with the CES-D, A Guide*]. Noordelijk Centrum voor Gezondheidsvraagstukken, Rijksuniversiteit Groningen.
Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for experimenters: Design, innovation, and discovery* (Vol. 2). Wiley Online Library.
Carlbring, P., Hagglund, M., Luthstrom, A., Dahlin, M., Kadowaki, A., Vernmark, K., et al. (2013). Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. *Journal of Affective Disorders, 148*, 331–337.
Christensen, H., Griffiths, K. M., & Farrer, L. (2009). Adherence in internet interventions for anxiety and depression. *Journal of Medical Internet Research, 11*, e13.
Collins, L. M., Murphy, S. A., Nair, V. N., & Strecher, V. J. (2005). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine, 30*, 65–73.
Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *American Journal of Preventive Medicine, 32*, S112–S118.
Cuijpers, P., & Smit, F. (2004). Subthreshold depression as a risk indicator for major depressive disorder: a systematic review of prospective studies. *Acta Psychiatrica Scandinavica, 109*, 325–331.
Cuijpers, P., van Straten, A., Warmerdam, L., & van Rooy, M. J. (2010). Recruiting participants for interventions to prevent the onset of depressive disorders: possible ways to increase participation rates. *BMC Health Services Research, 10*, 181.
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 1–38.
de Vries, H., & Brug, J. (1999). Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. *Patient Education and Counseling, 36*, 99–105.
Donkin, L., Christensen, H., Naismith, S. L., Neal, B., Hickie, I. B., & Glozier, N. (2011). A systematic review of the impact of adherence on the effectiveness of e-Therapies. *Journal of Medical Internet Research, 13*.
Fledderus, M., Bohlmeijer, E. T., Pieterse, M. E., & Schreurs, K. M. (2011). Acceptance and commitment therapy as guided self-help for psychological distress and positive mental health: a randomized controlled trial. *Psychological Medicine*, 1–11.
Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do*. Boston: Morgan Kaufmann Publishers.
Fry, J. P., & Neff, R. A. (2009). Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. *Journal of Medical Internet Research, 11*.
Furmark, T., Carlbring, P., Hedman, E., Sonnenstein, A., Clevberger, P., Bohman, B., et al. (2009). Guided and unguided self-help for social anxiety disorder: randomised controlled trial. *British Journal of Psychiatry, 195*, 440–447.
Glasgow, R. E. (2007). eHealth evaluation and dissemination research. *American Journal of Preventive Medicine, 32*, S119–S126.
Haringsma, R., Engels, G. I., Beekman, A. T. F., & Spinhoven, P. (2004). The criterion validity of the Center for Epidemiological Studies Depression scale (CES-D) in a sample of self-referred elders with depressive symptomatology. *International Journal of Geriatric Psychiatry, 19*, 558–563.
Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and commitment therapy: an experiential approach to behavior change*. New York: Guilford Press.
Hedman, E., Lindefors, N., Andersson, G., Andersson, E., Lekander, M., Ruck, C., et al. (2013). Predictors of outcome in internet-based cognitive behavior therapy for severe health anxiety. *Behaviour Research and Therapy, 51*, 711–717.
Hilvert-Bruce, Z., Rossouw, P. J., Wong, N., Sunderland, M., & Andrews, G. (2012). Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. *Behaviour Research and Therapy, 50*, 463–468.
Hurling, R., Fairley, B. W., & Dias, M. B. (2006). Internet-based exercise intervention systems: are more interactive designs better? *Psychology & Health, 21*, 757–772.
Johansson, R., & Andersson, G. (2012). Internet-based psychological treatments for depression. *Expert Review of Neurotherapeutics, 12*, 861–869. quiz 870.
Johansson, R., Sjoberg, E., Sjogren, M., Johnsson, E., Carlbring, P., Andersson, T., et al. (2012). Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. *PLoS ONE, 7*.
Jorm, A. F., & Griffiths, K. M. (2006). Population promotion of informal self-help strategies for early intervention against depression and anxiety. *Psychological Medicine, 36*, 3–6.
Kaltenthaler, E., Parry, G., Beverley, C., & Ferriter, M. (2008). Computerised cognitive-behavioural therapy for depression: systematic review. *British Journal of Psychiatry, 193*, 181–184.
Kelders, S. M. (2015). Involvement as a working mechanism for persuasive technology. In *Persuasive technology: 10th international conference, persuasive 2015, Chicago, IL, USA, June 3-5, 2015, Proceedings* (Vol. 9072, p. 3). Springer.
Kelders, S. M., Bohlmeijer, E. T., & van Gemert-Pijnen, J. E. (2013). Participants, usage, and use patterns of a web-based intervention for the prevention of depression within a randomized controlled trial. *Journal of Medical Internet Research, 15*(8).
Kelders, S. M., Kok, R. N., Ossebaard, H. C., & Van Gemert-Pijnen, J. E. (2012). Persuasive system design does matter: a systematic review of adherence to web-based interventions. *Journal of Medical Internet Research, 14*, e152.
Kelders, S. M., Pots, W. T., Oskam, M. J., Bohlmeijer, E. T., & van Gemert-Pijnen, J. E. (2013). Development of a web-based intervention for the indicated prevention of depression. *BMC Medical Informatics and Decision Making, 13*, 26.
Kelders, S. M., Van Gemert-Pijnen, J. E., Werkman, A., Nijland, N., & Seydel, E. R. (2011). Effectiveness of a Web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. *Journal of Medical Internet Research, 13*, e32.
Knutov, E., De Bra, P., & Pechenizkiy, M. (2009). AH 12 years later: a comprehensive survey of adaptive hypermedia methods and techniques. *New Review of Hypermedia and Multimedia, 15*, 5–38.
Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. *American Psychologist, 48*, 1181.
Morgan, A. J., Jorm, A. F., & Mackinnon, A. J. (2012). Email-based promotion of self-help for subthreshold depression: mood memos randomised controlled trial. *British Journal of Psychiatry, 200*, 412–418.
Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105.
Morrison, L. G., Yardley, L., Powell, J., & Michie, S. (2012). What design features are used in effective e-Health interventions? A review using techniques from critical interpretive synthesis. *Telemedicine and E-Health, 18*, 137–144.
Musiat, P., & Tarrier, N. (2014). Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. *Psychological Medicine, FirstView*, 1–14.
Noar, S. M., Benac, C. N., & Harris, M. S. (2007). Does tailoring matter? meta-analytic review of tailored print health Behavior change interventions. *Psychological Bulletin, 133*, 673–693.
Norman, G. J. (2008). Answering the "What works?" question in health behavior change. *American Journal of Preventive Medicine, 34*, 449–450.

Oinas-Kukkonen, H. (2010). Behavior change support systems: a research model and agenda. *Persuasive Technology, Proceedings, 6137*, 4–14.
Oinas-Kukkonen, H., & Harjumaa, M. (2009). Persuasive systems design: key issues, process model, and system features. *Communications of the Association for Information Systems, 24*, 28.
Olsson, I., Mykletun, A., & Dahl, A. A. (2005). The hospital anxiety and depression rating scale: a cross-sectional study of psychometrics and case finding abilities in general practice. *BMC Psychiatry, 5*.
Pots, W. T. M., Fledderus, M., Meulenbeek, P. A. M., Ten Klooster, P. M., Schreurs, K. M. G., & Bohlmeijer, E. T. (2015). Acceptance and commitment therapy as a web-based intervention for depressive symptomatology: randomised controlled trial. *British Journal of Psychiatry* (in press).
Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.
Ritterband, L. M., Cox, D. J., Gordon, T. L., Borowitz, S. M., Kovatchev, B. P., Walker, L. S., et al. (2006). Examining the added value of audio, graphics, and interactivity in an internet intervention for pediatric encopresis. *Children's Health Care, 35*, 47–59.
Spek, V., Cuijpers, P., Nyklicek, I., Riper, H., Keyzer, J., & Pop, V. (2007). Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychological Medicine, 37*, 319–328.
Spinhoven, P., Ormel, J., Sloekers, P. P. A., Kempen, G. I. J. M., Speckens, A. E. M., & VanHemert, A. M. (1997). A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychological Medicine, 27*, 363–370.
Strecher, V. J., McClure, J. B., Alexander, G. L., Chakraborty, B., Nair, V. N., Konkel, J. M., et al. (2008). Web-based smoking-cessation programs: results of a randomized trial. *American Journal of Preventive Medicine, 34*, 373–381.
Talbot, F. (2012). Client contact in self-help therapy for anxiety and depression: necessary but can take a variety of forms beside therapist contact. *Behaviour Change, 29*, 63–76.
Tauer, J. M., & Harackiewicz, J. M. (1999). Winning isn't everything: competition, achievement orientation, and intrinsic motivation. *Journal of Experimental Social Psychology, 35*, 209–238.
Titov, N., Andrews, G., Choi, I., Schwencke, G., & Johnston, L. (2009). Randomized controlled trial of web-based treatment of social phobia without clinician guidance. *Australian and New Zealand Journal of Psychiatry, 43*, 913–919.
Titov, N., Andrews, G., Davies, M., McIntyre, K., Robinson, E., & Solley, K. (2010). Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. *PLoS ONE, 5*, e10939.
Velsen, L. S. (2011). *User-centered design for personalization*. University of Twente.
Webb, T. L., Joseph, J., Yardley, L., & Michie, S. (2010). Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *Journal of Medical Internet Research, 12*, e4.
Zaichkowsky, J. L. (1994). The personal involvement inventory - reduction, revision, and application to advertising. *Journal of Advertising, 23*, 59–70.
Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*, 361–370.