

A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods



Kien Hoa Ly^{a,*}, Ann-Marie Ly^b, Gerhard Andersson^{a,c}

^a Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

^b Department of Psychology, Mittuniversitetet, Östersund, Sweden

^c Department of Clinical Neuroscience, Center for Psychiatry Research, Karolinska Institutet, Stockholm, Sweden

ABSTRACT

Fully automated self-help interventions can serve as highly cost-effective mental health promotion tools for massive amounts of people. However, these interventions are often characterised by poor adherence. One way to address this problem is to mimic therapy support by a conversational agent. The objectives of this study were to assess the effectiveness and adherence of a smartphone app, delivering strategies used in positive psychology and CBT interventions via an automated chatbot (Shim) for a non-clinical population — as well as to explore participants' views and experiences of interacting with this chatbot. A total of 28 participants were randomized to either receive the chatbot intervention ($n = 14$) or to a wait-list control group ($n = 14$). Findings revealed that participants who adhered to the intervention ($n = 13$) showed significant interaction effects of group and time on psychological well-being (FS) and perceived stress (PSS-10) compared to the wait-list control group, with small to large between effect sizes (Cohen's d range 0.14–1.06). Also, the participants showed high engagement during the 2-week long intervention, with an average open app ratio of 17.71 times for the whole period. This is higher compared to other studies on fully automated interventions claiming to be highly engaging, such as Woebot and the Panoply app. The qualitative data revealed sub-themes which, to our knowledge, have not been found previously, such as the moderating format of the chatbot. The results of this study, in particular the good adherence rate, validated the usefulness of replicating this study in the future with a larger sample size and an active control group. This is important, as the search for fully automated, yet highly engaging and effective digital self-help interventions for promoting mental health is crucial for the public health.

1. Introduction

Over nearly two decades, researchers have gathered evidence to support the notion that rather simple cognitive and behavioural strategies, designed to mirror the thoughts and behaviours of naturally happy people, can help increase well-being, reduce negative symptoms and improve happiness (Bolier et al., 2013; Sin and Lyubomirsky, 2009; Weiss et al., 2016). By combining elements of positive psychology and cognitive behaviour therapy (CBT), this approach aims to increase the individual's engagement in meaningful and value-congruent behaviour, as well as in cognition focusing on the meaningful and valuable aspects of his or her life. This in turn is shown to translate into higher psychological well-being, increased levels of global life satisfaction and reduced stress levels (Bolier et al., 2013; Layous and Lyubomirsky, 2014). In addition to this, findings show that individuals who are experiencing high levels of psychological well-being tend to have the lowest number of health care visits, hospital days and drug

consumption (Keyes and Grzywacz, 2005). Thus, these interventions can save societies enormous amount of money indirectly.

One of the first meta-analytical reviews to explore the effects of this approach combined the results of 49 studies to conclude that these types of activities can significantly increase well-being (mean Cohen's $d = 0.61$), and decrease depressive symptoms (mean Cohen's $d = 0.65$) (Sin and Lyubomirsky, 2009). However, this meta-analysis had some crucial limitations, such as including both randomized and quasi-experimental studies. A new meta-analysis, entailing 39 studies, was conducted a few years later — and included a different set of studies, where the design quality was assured using randomized controlled trials only. The mean effect sizes from a combined analysis of these 39 studies showed $d = 0.34$ on subjective well-being, $d = 0.20$ on psychological well-being, and $d = 0.23$ on depression (Bolier et al., 2013), which is about 0.3 points lower than the effect sizes estimated in the first meta-analysis. The authors attributed this difference to the more rigorous studies included in their review (Bolier et al., 2013). In a third

* Corresponding author.

E-mail addresses: kien.hoa.ly@liu.se (K.H. Ly), anly1376@student.uu.se (A.-M. Ly), gerhard.andersson@liu.se (G. Andersson).

<http://dx.doi.org/10.1016/j.invent.2017.10.002>

Received 14 August 2017; Received in revised form 2 October 2017; Accepted 3 October 2017

Available online 10 October 2017

2214-7829/ © 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

meta-analysis, the effects of behavioural interventions on psychological well-being were examined. Based on 27 studies, a moderate effect of $d = 0.44$ was found (Weiss et al., 2016), which is somewhat lower than the mean effect sizes of well-being in the Sin and Lyubomirsky (2009) meta-analysis ($d = 0.61$) and somewhat higher than the effect sizes of psychological well-being in the Bolier et al. (2013) study ($d = 0.20$). Moreover, in Bolier et al. (2013) and Weiss et al. (2016) reviews, long-term effects were examined. Both studies found small, but significant effect sizes at follow-up (up to ten months) on well-being (Bolier et al., 2013; Weiss et al., 2016). However, these follow-up results should be interpreted with caution because of the small number of studies and the high attrition rates at follow-up.

In conclusion, it appears as activities and exercises that promote positive feelings, thoughts, and/or behaviours, such as expressing gratitude, practicing kindness (Parks et al., 2012a; Dunn et al., 2008; Sheldon et al., 2012), engaging in enjoyable activities (Fordyce, 1977) and replaying positive experiences (Fava et al., 1998), can have a positive impact on well-being and negative symptoms for a broad range of people, both short and long-term — although additional high-quality peer-reviewed studies are needed to strengthen the evidence-base for these types of activities (Bolier et al., 2013).

Several characteristics of the interventions influenced the effects of the outcome, with a couple of them being found across all three meta-analyses: Face-to-face on an individual basis was most effective, followed by group-interventions — and lastly, self-help (Bolier et al., 2013; Sin and Lyubomirsky, 2009; Weiss et al., 2016). Also, larger effects were found for clinical groups, although non-clinical samples respond well to these type of interventions (Lewandowski, 2009; Schutte et al., 2012; Smyth, 1998). This finding could be explained by the fact that clinical populations have more room for improvement and therefore show larger positive changes (Bolier et al., 2013; Sin and Lyubomirsky, 2009; Weiss et al., 2016). In spite of these findings, it has been argued that delivering these types of interventions as self-help to a broad population is the most beneficial from a public health perspective (Bolier et al., 2013), since they are relatively brief, easy to administer — and at the same time demonstrate at least small significant effect sizes (Sin and Lyubomirsky, 2009). This combination means that these self-help interventions can serve as highly cost-effective mental health promotion tools for massive amounts of people, some of which may not otherwise be reached by mental health care (Bergsma, 2008; den Boer et al., 2004; Parks et al., 2012b). Thus, they can have a major impact on populations' well-being (Huppert, 2009). In addition, the self-help format suits the goals of positive psychology very well, since the idea behind these activities and interventions is to avoid making any major shifts to people's current life situations, thus being an integrated part of people's everyday lives — and can be used by a broad population, spanning from clinical to healthy people (Sin and Lyubomirsky, 2009).

Despite the many proven benefits of this type of self-help interventions, there are also significant problems related to the approach. The most central one, according to current literature, is its low adherence (Christensen et al., 2009; Schueller, 2010). Most notably, this is a major problem for fully automated self-help interventions, while interventions entailing some element of therapist support have at times overcome this issue (Andersson, 2016). Enhancing adherence of fully automated self-help interventions could therefore be a major factor in improving effectiveness. Given this, a critical question is how to involve and keep people engaged in well-being-enhancing activities. A hypothesized reason for this lack of adherence is, in line with the above, the loss of the human interactional quality of in-person interventions. For example, certain therapeutic process factors such as accountability may be more salient in traditional face-to-face treatments, compared to purely digital health interventions. One way to address this problem is to augment self-help interventions with external support from clinicians or coaches. While this approach holds promise in for example self-guided depression treatment (Mohr et al., 2013), its ability to scale widely is limited due to costs, availability of coaches, and scheduling

logistics (Morris et al., 2015).

Another approach to this problem would be to mimic human dialogues via a conversational agent. Conversational agents, such as Siri (Apple), S Voice (Samsung), and Cortana (Microsoft), are smartphone-based computer programs designed to respond to users in natural language, thereby mimicking conversations between people. In a recent study, it was demonstrated that a text-based conversational agent could significantly reduce symptoms of depression and anxiety among a non-clinical college student population, compared to an information-only control group (Fitzpatrick et al., 2017). Although the results need to be replicated and the study failed to show any significant differences in adherence between the two groups, the results are promising and show that a conversational agent can be a powerful medium through which individuals receive mental health interventions. Also, a systematic review has shown that interventions based on synchronous written conversations (i.e. "chats") can have significant and sustained effects on mental health outcomes as compared to wait-list condition — and equivalent to face-to-face therapy and telephone counselling (Hoermann et al., 2017). Continuing exploring if this approach is a way to develop effective, scalable and highly engaging interventions to promote mental health for people in a broad range of populations, is not only legitimate, but crucial in order to have a major impact on populations' well-being.

The objective of this study was to assess the effectiveness and adherence of delivering strategies, used in positive psychology and CBT interventions (third wave of CBT) such as expressing gratitude and replaying positive experiences, in a conversational interface via an automated smartphone-based chatbot (a computer program which conducts a conversation via auditory or textual methods). A second aim was to explore participants' views and experiences of interacting with this chatbot. The study was designed as a randomized controlled study in a non-clinical population, and compared outcomes from two weeks' usage of a positive psychology-oriented conversational agent (the iPhone application *Shim*), against a wait-list control group. We hypothesized that conversation with a therapeutic process-oriented conversational agent would lead to increased levels of subjective, social and psychological well-being, as well as decreased levels of perceived stress, compared to the wait-list control group.

2. Methods

2.1. Design

This was a pilot randomized controlled trial, conducted in Sweden, comparing a smartphone application (app), built as a conversational agent to deliver a positive psychology and CBT intervention ($n = 14$) against a wait-list control group ($n = 14$) for a non-clinical population.

2.2. Ethics statement

Since this pilot trial involved a non-clinical population, it was considered exempt from registration in a public trials registry. The institutional review board approved the study protocol. Written informed consent was obtained from all participants electronically before the study started. Considerations were given to whether the intervention might be harmful to administer to a non-clinical population. For example, if a participant would come to realize that she lacks meaning in her life by these types of reflections. However, since the literature is rather showing the opposite; non-clinical populations benefit from these type of interventions (Lewandowski, 2009; Schutte et al., 2012; Smyth, 1998), we concluded that the possible health benefits would exceed the negative ones.

2.3. Recruitment and selection

It was a self-selected sample, recruited from various Swedish

universities, as well as from a website dealing with positive psychology and via social media channels (Facebook, Twitter). Those who were interested were directed to a web page with information about the study, the app being tested and how to participate. From the web page, the participants were able to fill out an online screening assessment, which was necessary to be completed in order to be included in the study. Except from provision of the Iphone app, no inducement for participation was offered.

2.4. Participants

Although we did not screen for major psychiatric conditions, the study was advertised as a non-clinical trial with an opportunity to try a new conversational smartphone app directed towards positive psychology and CBT. In order to be included, participants had to fulfil the following criteria: a) at least 18 years of age; b) not participate in any psychological treatment; c) not consume any psychopharmacological drugs — or if doing so; be on a fixed dose for more than one month; d) have continuous access to an Iphone.

In total, 36 individuals initially expressed interest in the study. One person did not complete all the questions in the online screening, four people were excluded because of an ongoing psychological treatment and one person did not have the right technical requirements in her Iphone. In addition to this, two participants decided to leave the study before it had started. Finally, 28 participants were included in the data analyses. The flow of participants through the study is shown in Fig. 1. Among the randomized participants, there were 53.6% women ($n = 15$) and 46.4% men ($n = 13$). The mean age was 26.2 years ($SD = 7.2$) ranging from 20 to 49 years. See Table 1 for additional

Table 1

Demographic description of the participants.

		Shim group ($n = 14$)	Control group ($n = 14$)	Interviewed ($n = 9$)
Age	Mean (SD)	21.1 (8.8)	25.4 (5.3)	28.8 (9.7)
	Min-max	20–49	20–39	20–49
Gender, n (%)	Female	7 (50)	8 (57)	5 (56)
	Male	7 (50)	6 (43)	4 (44)
Occupation, n (%)	Working (full-time)	5 (36)	5 (36)	5 (56)
	Student	9 (64)	9 (64)	4 (44)
Marital status, (%)	Married/cohabitant	4 (29)	4 (29)	3 (33)
	Partner but not living together	2 (14)	5 (36)	1 (11)
	Single/divorced	8 (57)	5 (36)	5 (56)

demographical data.

The interviewed participants, providing qualitative data, were nine individuals from the intervention group who agreed to also participate in this part of the study. The demographics of these individuals were 44.4% women ($n = 4$) and 55.6% men ($n = 5$). The mean age was 28.8 years ($SD = 9.7$) ranging from 20 to 49 years.

2.5. Outcome measures

The outcome measures were The Flourishing Scale (FS), The Satisfaction With Life Scale (SWLS), and The Perceived Stress Scale (PSS-10).

The FS (previously The Psychological Well-Being scale) is a brief 8-item summary measure of the respondent's self-perceived success in important areas such as relationships, self-esteem, purpose, and optimism (Diener et al., 2009). The scale provides a single psychological well-being score and is widely used in well-being intervention studies because of its brevity, simplicity and comprehensiveness (Schotanus-Dijkstra et al., 2016). FS has shown adequate to excellent reliability with Cronbach's alpha values, ranging from 0.78 to 0.95, as well as moderate to strong positive correlations for overall psychological well-being — and moderate to strong negative correlations for depression, anxiety and stress (Diener et al., 2010).

The SWLS is a 5-item self-report, 7-point scale concerning subjective well-being (Diener et al., 1985), which is assessed by measuring cognitive self-judgment about satisfaction with one's life. SWLS has shown strong internal reliability (Cronbach's Alpha = 0.87) and good temporal stability (test-retest correlation $r = 0.82$) (Diener et al., 1985).

The PSS-10 measures the perception of stress, i.e., the degree to which situations are appraised as stressful, by asking respondents to rate the frequency of their thoughts and feelings related to situations that occurred in the recent time (Cohen et al., 1994). It consists of 10 items rated on a five-point Likert-type scale, ranging from “Never” to “Very Often”. PSS is one of the most widely used psychological instruments (Cohen et al., 1994).

2.6. Administration format of self-report measures

We used an online platform to administer the FS, the SWLS and the PSS-10. Previous psychometric research has validated internet-administration of self-rating scales in a various number of interventions (Carlborg et al., 2007; Hedman et al., 2010; Holländare et al., 2010).

2.7. Procedure

For those participants included in the study, the results from the

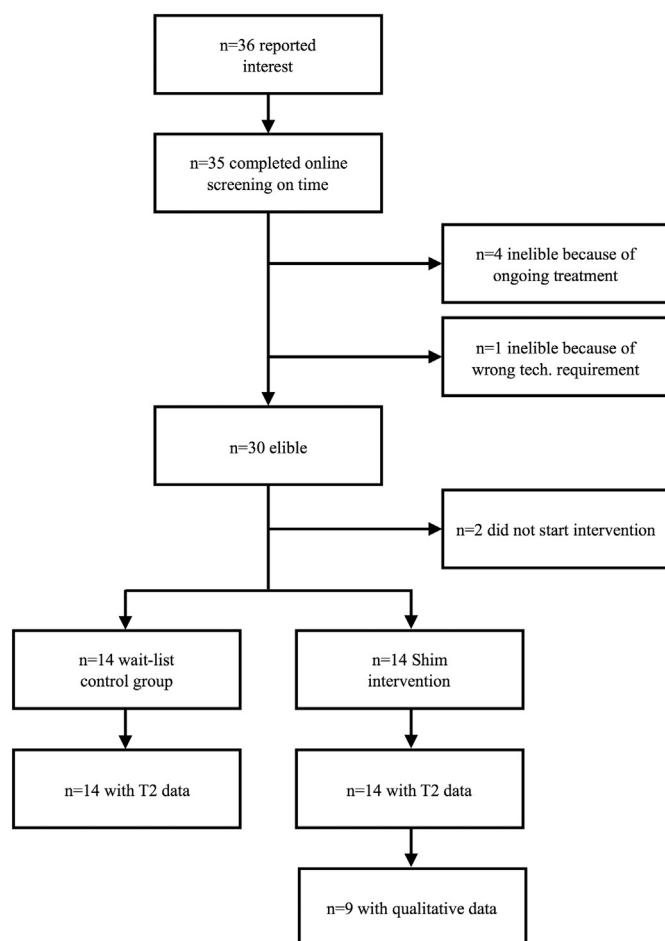


Fig. 1. Participant flow chart.

online screening were used as pre-treatment assessment. After the recruitment, participants were allocated using an online randomization tool (www.random.org) in a 1:1 ratio. The randomization procedure was handled by an independent person who was separate from the staff conducting the study. Post-test was assessed directly after the intervention.

The qualitative data were gathered at post-measurement through semi-structured telephone interviews, which lasted 20 to 30 min each. An interview guide was prepared with a small set of questions targeting participants' good and bad experiences of the intervention. In order to gain as rich information as possible, these questions were open-ended with subsequent follow-up questions. The interview guide was used rather as an inspiration than as a mandatory instrument.

2.8. The intervention

Shim is a fully automated conversational agent, built as a smartphone app. The user interface (UI) of Shim is similar to a text messaging app. The conversations in Shim are centred around insights, strategies and activities related to the field of positive psychology. These include (but are not limited) to expressing gratitude, practicing kindness, engaging in enjoyable activities and replaying positive experiences. Also, components from the third wave of CBT are included in the strategies taught by Shim, such as present moment awareness, valued directions and committed actions. The goal of the conversations in Shim is to help the user reflect upon, learn and practice these small strategies and behaviours. One example of what Shim asks the user to do is imagining being present at one's own 80's birthday. As the next step, the user is asked to reflect upon what she believes she would value the most in life – and what steps she could take to help her live a more value-based life. This exercise is similar to the funeral exercise described in for e.g. Dahl and Lundgren (2006). While this has been accomplished in other previous medium, such as self-help books, the idea behind Shim is to do this in a more natural way via a friendly dialogue, and as part of the user's everyday life. Also, the conversational agent medium made it possible to involve therapeutic process-oriented features in Shim, such as empathic responses based on user's mood, tailored content based on user's previous inputs, daily check-ins to create a sense of accountability, and weekly summaries in the end of each week containing all thoughts, reflections and activities accomplished by the user.

The dialogues in Shim have been pre-written by professionals with training in psychology. Each dialogue can be represented as a tree-graph with one or multiple starting points and one or multiple closing points. The user replies Shim mostly via two types of statements: (1) comments, requiring inputs from the user such as free text, (2) picking elements either from a list or from a fixed set of reply options. With language pattern matching and keyword spotting, as well as conditional expressions, Shim gives adequate responses to the user's statements.

Shim is most often the main driver of the conversations. When a user opens the app, Shim starts the dialogue, which is determined by an algorithm that takes into account external (e.g. what time of day, what day of week) and internal information (e.g. data about the user from previous conversations). The conversations are often questions about both small everyday things, such as to whom the user has felt gratitude towards lately – as well as bigger and deeper things, such as what strengths the user feels he/she possesses. The user replies to Shim mainly by either choosing between a few pre-defined answers or via free text. From time to time, Shim delivers insights and research findings from the positive psychology domain, such as the benefits of expressing gratitude. In addition to this, Shim also suggests small activities to do, such as getting in touch with a friend the user has not been in touch with for a while. Each time the user ends the conversation with Shim, the database for the current user is updated in line with the information provided by the user. This lets Shim grow its data about the user over time, as she talks to the app.

In summary, Shim comprises a simple but familiar messaging UI, a

technical solution based on a large set of pre-written dialogues, a rule system and a database management system to classify and store users' input to produce an interactive positive psychology and CBT intervention.

2.9. Data analysis

There is little published guidance with regards to how many participants should be included in a pilot RCT. However, it has been suggested that smaller sample sizes in the range of 10–20 participants per group can be sufficient (Hertzog, 2008; Kieser and Wassmer, 1996). A sample size calculation showed that a total number of 26 participants was required to achieve a power of 0.80 to detect a between-group effect size of $d = 0.60$ with the less liberal alpha level of 0.1, which can be used in pilot studies. All analyses were performed using SPSS 24 (IBM Corp. Released 2016. IBM SPSS Statistics for Macintosh, Version 24.0. Armonk, NY: IBM Corp.). Independent t -tests and χ^2 -tests were used to test for group differences in demographics and pre-treatment data. Differences between the intervention group and the control group were analysed using mixed effects models with Maximum Likelihood estimation. Also, random intercept models were selected for all measures. The differences between the two groups were investigated by modelling interaction effects of group and time. Between-group differences at post-treatment were analysed using independent t -tests. Within- and between-group effect sizes (Cohen's d) were calculated by dividing the differences in means by the pooled standard deviations. The mentioned analyses were conducted in accordance with the intention-to-treat (ITT) principle. In addition to this, a completer analysis was conducted with only participants who adhered to the intervention.

The interview data were processed with the aid of thematic analysis (Braun and Clarke, 2006) to detect themes that facilitate understanding of the data. The analysis was driven by empirical findings, i.e. an inductive process (Hayes, 2000). The first step of analysis was the transcription of the interviews, followed by several thorough readings in order to identify patterns in the data, reflecting common themes in the participants' experiences of the intervention. All text that referred to participants' positive or negative experiences was coded. A list of 12 keywords or subthemes were derived from this data. The subthemes are listed in Table 3 below.

When all interviews had been coded, these codes and their corresponding quotes were assembled in a document. Preliminary themes were then identified, and the codes were grouped under these themes. This procedure is in line with how thematic analysis is described in literature (Braun and Clarke, 2006).

3. Results

The two groups did not differ significantly on any of the measures at pre-treatment ($t(26) = -0.86$ to -0.21 , $p = 0.40$ to 0.84). Also, there were no significant differences in demographic characteristics between the groups ($\chi^2(1) = 0.00$ to 3.12 , $p = 0.37$ to 0.65). See Table 1 for demographical data and Table 2 for all outcome measurements at pre- and post-treatment.

3.1. Attrition and adherence

The typical web-based intervention is usually meant to be used once a week (Kelders et al., 2012). Also, our research group has previously defined adherence to a web-based treatment as the number of weekly reflections the participants complete, with a completer of the treatment to make at least 8 reflections during an 8-week period (Ly et al., 2014). Since Shim is a smartphone intervention and used differently than a web-based treatment, we defined being a completer of the intervention as completing at least 14 reflections over the course of 14 days, as well as not being inactive for 7 or more days in a row. Only one participant did not fulfil these criteria. We made a separate completer analysis with

Table 2Means, SDs and effect sizes (Cohen's *d*) for measures of psychological well-being, perceived stress and subjective well-being.

Outcome measures		Mean (SD)		Effect size, <i>d</i>	
		Pre-treatment	Post-treatment	Between group, post	Within-group, pre to post
<i>ITT analysis</i>					
FS					
<i>n</i> = 14	Shim group	44.43 (5.9)	45.14 (6.0)	0.01	0.12
<i>n</i> = 14	Control group	46.14 (4.7)	45.07 (5.7)		− 0.21
PSS-10					
<i>n</i> = 14	Shim group	15.36 (5.2)	12.93 (5.2)	0.91	0.49
<i>n</i> = 14	Control group	16.86 (5.0)	17.14 (4.4)		− 0.06
SWLS					
<i>n</i> = 14	Shim group	25.50 (5.2)	26.79 (6.0)	0.15	0.24
<i>n</i> = 14	Control group	25.86 (3.9)	26.00 (4.6)		0.03
<i>Completer analysis</i>					
FS					
<i>n</i> = 13	Shim group	44.38 (6.1)	45.85 (5.6)	0.14*	0.26
<i>n</i> = 14	Control group	46.14 (4.7)	45.07 (5.7)		− 0.21
PSS-10					
<i>n</i> = 13	Shim group	16.23 (4.2)	12.38 (5.0)	1.05*	0.87*
<i>n</i> = 14	Control group	16.86 (5.0)	17.14 (4.4)		− 0.06
SWLS					
<i>n</i> = 13	Shim group	25.62 (5.4)	27.69 (5.2)	0.36	0.41*
<i>n</i> = 14	Control group	25.86 (3.9)	26.00 (4.6)		0.03

Abbreviations: FS: The Flourishing Scale; PSS-10: The Perceived Stress Scale; SWLS: The Satisfaction With Life Scale.

* $p < 0.05$.

only participants who adhered to the intervention.

In total, 11 (78.6%) out of the 14 participants were active at least 50% of the days. The average active days was 8.21 ($SD = 3.0$) meaning that the average participant was active for more than half of the intervention's 14 days. In addition to this, we measured how many times per day participants opened the app to have a conversation with Shim (1 conversation contained in average 3 reflections). The variation was large, ranging from 12 times for a day down to 0 times. The most active participant had an average opening app ratio of 4.43 times/day and the least active participant 0.29 times/day. The average app open app ratio was 1.27 times/day across all participants. **The average open app ratio for the whole 2-week period** was 17.71 times ($SD = 15.7$). All of the 28 randomized participants provided post-measurement data.

3.2. Outcome measures

No significant interaction effects of group and time were found between the intervention group and the control group on any of the measurements (FS: ($F_{1, 26} = 1.76$, $p = 0.20$); SWLS: ($F_{1, 26} = 0.67$, $p = 0.42$); PSS-10: ($F_{1, 26} = 1.23$, $p = 0.28$)).

3.3. Completer analysis

When only analysing participants, who had adhered to the intervention (i.e. active at least 25% of the days and not being inactive for 7 days or more in a row), significant interaction effects of group and time on the FS and the PSS-10 were found between the intervention group and the control group: FS: ($F_{1, 27} = 5.12$, $p = 0.032$); PSS-10: ($F_{1, 27} = 4.30$, $p = 0.048$). Between group effect size for the FS was below small; $d = 0.14$, which can be attributed to the much lower mean score on FS at pre-measurement for the intervention group (44.38) compared to the control group (46.14). For the PSS-10, the between group effect size was large $d = 1.06$. No significant interaction effects of group and time were found on the SWLS: ($F_{1, 27} = 2.83$, $p = 0.10$). In addition to this, large to medium significant within-group effect sizes were found on the PSS-10: $t(12) = 2.22$; $p = 0.046$; $d = 0.87$; and the SWLS: $t(12) = -2.25$; $p = 0.044$; $d = 0.41$.

3.4. Qualitative data

The qualitative data generated insights into what benefits and opportunities, as well as challenges there might be with an automated smartphone-based chatbot, targeting a non-clinical population. Three main themes with attached subthemes were found. The themes were **Content** with the subthemes *activation*, *learning*, *reflection*, *repetitiveness* and *shallowness*; **Medium** with the subthemes *routine*, *moderator*, *availability* and *lack of clarity*; and lastly **Functionalities** with the subthemes *weekly summary*, *lack of notifications* and *restricted UI*.

3.4.1. Content

It was clear that the content was a central part of the app. However, it was a mix of what parts of the content participants appreciated the most. Some participants highlighted that they liked the learning part the most, when the app taught them positive psychology principles:

It was so nice to learn things from Shim.

(Participant 4)

Other participants said their best experience with Shim was when the app encouraged them to express gratitude or show appreciation to a close one:

The best experience with Shim was when I got nudged to send a text message after I had reflected on how the day was. When I sent a nice text message to someone I got very happy about it. And when I got a nice response back, that was such a wow-feeling!

(Participant 2)

The above quote shows that the participant's behaviour also triggered a behaviour in the recipient, which added an even more positive feeling to the experience.

A third distinct positive subtheme related to the content and what participants appreciated the most, was that the app helped to reflect upon important aspect of life:

In general, the best thing with Shim was that it made me reflect on important things in life.

(Participant 5)

One participant also highlighted the mix of learning and reflections:

Table 3

Participants' themes presented as main themes, subthemes and their quotes.

Participant's quotes	Subthemes	Main themes
<i>The best thing with Shim is that it made me do things I would not otherwise have done, such as positive things and show appreciation and gratitude to others. (Participant 1)</i>	Activation	Content
<i>My best experience with Shim was when I got nudged to send a text message after I had reflected on how the day was. When I sent a nice text message to someone I got very happy about it. And when I got a nice response back, that was such a wow-feeling! (Participant 2)</i>		
<i>Something particular that I liked was that Shim encouraged me to keep in contact with my loved ones. (Participant 3)</i>		
<i>A wow experience I had was when I got nudged by Shim to text message a friend and tell the person why I appreciate his friendship. That was an eye-opener for me. I never do this otherwise. (Participant 6)</i>		
<i>Shim is very good when first guiding me to think in a more constructive way and about positive things — and then tries to change my behaviours. (Participant 8)</i>		
<i>It was so nice to learn things from Shim. (Participant 4)</i>	Learning	
<i>I liked the blend between learning stuff from Shim and reflecting and giving things back to Shim. (Participant 1)</i>		
<i>In general, the best thing with Shim was that it made me reflect on important things in life. (Participant 5)</i>	Reflection	
<i>There's no app that has made it so easy to reflect on things in life as Shim. You can get something out of your time in a better way than just checking Facebook or some webpage by reflecting upon important things. (Participant 6)</i>		
<i>It made me think and reflect. (Participant 1)</i>		
<i>After a while it becomes the same thing (Participant 5)</i>		
<i>The only negative feeling I got when using Shim was when Shim asked me the same thing again. (Participant 1)</i>	Repetitiveness	
<i>After a couple of times using Shim, I felt I didn't put as much energy to reflect upon the questions. This was a consequence of seeing the same questions again. (Participant 3)</i>		
<i>The same question about relationship to another person came quite a lot and in a mechanical way. Shim asked this in the exact same way for all my five friends. A real person would not ask the exact same question about all my friends. (Participant 4)</i>		
<i>The next day, Shim proposed the same message for the same friend. It really felt like texting with a machine when this happened. (Participant 6)</i>		
<i>I felt Shim could talk to me about stuff on the surface, but not really deepened the relationship. At that point, I felt Shim couldn't take any more steps in our relationship. I kind of felt: "this is as far we can get". (Participant 5)</i>	Shallowness	
<i>I wanted to know if Shim would take our relationship to a deeper level. During these 14 days, I must say it didn't develop that much. (Participant 7)</i>		
<i>When I talked to Shim about my grandparents, Shim didn't ask and understand they had passed away. I felt disappointed when this happened, and felt that this relationship will have boundaries, it won't get too deepened. (Participant 8)</i>		
<i>It became a small routine for me. I talked to Shim during evenings a couple of times as well, but mornings worked better for me — I knew I had a time slot by myself which was perfect for talking to Shim. (Participant 5)</i>	Routine	Medium
<i>I made a routine of it. I did it because I thought it was important to me. I always used Shim before going to bed, and mostly one time a day. (Participant 2)</i>		
<i>It became a nice routine. It was easy to get started with this routine. Normally, I don't use any apps on the bus, only listening to music, but this was a nice thing to do — I got my own little time with Shim. (Participant 4)</i>		
<i>Shim became a diary, a forum about and for yourself. I liked the diary aspect of Shim, and to be asked about things I did today which are worth remembering. (Participant 6)</i>		
<i>The unique and positive thing with Shim is that I could get in contact with Shim whenever I wanted to. (Participant 8)</i>	Availability	
<i>A big advantage was that I could pick up Shim whenever I wanted. For me, knowing I had it with me all the time, and that I got to talk to it every day made me get a nice, positive feeling. (Participant 1)</i>		
<i>That it was so easily accessible was really good. I could just pick Shim up during a break a start messaging with it. (Participant 6)</i>		
<i>I feel I get help from outside, like a person is guiding me and asking things I wouldn't have thought of myself (Participant 5)</i>	Moderator	
<i>It was like someone has made the work for me, deciding what questions to reflect upon (Participant 2)</i>		
<i>Shim functions as a moderator of what questions to focus on. (Participant 6)</i>		
<i>Shim needs a much clearer goal. I think this is one of the biggest problem now with Shim that the user has to "decode" the app. You don't understand what Shim is about. (Participant 9)</i>	Lack of clarity	
<i>I liked the weekly summary. It made me see much clearer what's important and it reminded me of the people I had contacted. (Participant 3)</i>	Weekly summary	Functionalities
<i>The surprising effect of the summaries were fantastic (Participant 7)</i>		
<i>What's really missing with Shim right now is the red circle on Facebook — the signal of a new notification. (Participant 9)</i>	Lack of notifications	
<i>One thing that felt a bit negative was that I couldn't write my own characteristics of people. But when I think of it, I'm not sure what I missed (Participant 2)</i>	Restricted UI	

I liked the blend between learning stuff from Shim and reflecting and giving things back to Shim.

(Participant 1)

The most evident limitation in the service at the time for the study, was the repetitiveness of the content. This was highlighted by a majority of the interviewed participants. An interesting finding related to this subtheme was that two of the participants expressed that the repetitiveness made the conversational agent feel less humanlike, thus indicating they initially had the perception of the app as somewhat a living character:

A real person would not ask the exact same question about all my friends.

(Participant 4)

It really felt like texting with a machine when this happened.

(Participant 6)

Another closely related subtheme was the app's limitation in taking the relationship to a deeper level:

I felt Shim could talk to me about stuff on the surface, but not really deepened the relationship. At that point, I felt Shim couldn't take any more steps in our relationship. I kind of felt: "this is as far we can get".

(Participant 5)

A more concrete example of a situation when the app lacked the ability to go deeper was described by this participant:

When I talked to Shim about my grandparents, Shim didn't ask and understand they had passed away. I felt disappointed when this happened, and felt that this relationship will have boundaries, it won't get too deepened.

(Participant 8)

Again, these quotes have an underlining tone of Shim being perceived as a living character, and even making one participant disappointed for not following up on some certain questions.

3.4.2. Medium

Another viable part in Shim was the medium. One salient subtheme related to the medium was that using the app became a routine for some participants:

It became a nice routine. It was easy to get started with this routine. Normally, I don't use any apps on the bus, only listening to music, but this was a nice thing to do — I got my own little time with Shim.
(Participant 4)

One explanation why Shim became a routine for some participants could be because they enjoyed the part of the app when it asked about participant's day. It could also be related to the next subtheme; availability. One participant highlighted that since Shim was always accessible, it was convenient to talk to Shim every day, which in turn gave her a nice feeling:

A big advantage was that I could pick up Shim whenever I wanted. For me, knowing I had it with me all the time, and that I got to talk to it every day made me get a nice, positive feeling.
(Participant 1)

As described above, Shim is the main driver of the conversation, which three of the participants explicitly highlighted as a positive aspect of the app:

It was like someone has made the work for me, deciding what questions to reflect upon
(Participant 2)

The above participant clearly perceived it as the app provided a value by driving the conversation and giving the participant relevant questions to reflection on. The same value was noticed by another participant:

I feel I get help from outside, like a person is guiding me and asking things I wouldn't have thought of myself
(Participant 5)

Once more, this quote highlights the app of being seen like a living character, who was communicating with the participant. On the other hand, this new type of medium for delivering psychological self-help has at least one big challenge, namely clarity of aim and goals. This was emphasized by one of the participant:

Shim needs a much clearer goal. I think this is one of the biggest problem now with Shim that the user has to “decode” the app. You don't understand what Shim is about.
(Participant 9)

3.4.3. Functionalities

The last theme, functionalities, included quotes and subthemes related to functionalities in the app which participants either highlighted as valuable or as something which was currently missed in Shim. One example of the latter was the lack of notifications to remind the participant of the app's existence:

What's really missing with Shim right now is the red circle on Facebook — the signal of a new notification.
(Participant 9)

This could indicate that the participant wanted to use the app, but had not yet made a routine of using it. Two participants highlighted the weekly summaries as something valuable. One of them described the functionality in the following way:

I liked the weekly summary. It made me see much clearer what's important and it reminded me of the people I had contacted.
(Participant 3)

4. Discussion

4.1. Main findings

The objective of this study was to assess the effectiveness and adherence of delivering strategies, used in positive psychology and CBT interventions (third wave of CBT) in a conversational interface via an automated smartphone-based chatbot. When analysing the whole sample, the result showed that the intervention did not differ significantly from the wait-list control group on any of the outcome measures. However, when only including the participants who had adhered to the intervention, we found significant interaction effects of group and time on measurement on psychological well-being (FS) as well as perceived stress (PSS-10), with small to large between effect sizes.

In general, the participants showed high engagement during the 2-week long intervention, with 78.6% of the participants being active 50% or more of the days. In addition to this, the average open app ratio for the whole 2-week period was 17.71 times. Compared to other studies on fully automated interventions claiming to be highly engaging, such as *Woebot* (in average 12.14 times during 2 weeks) (Fitzpatrick et al., 2017) and the *Panoply* app (in average 21 times during 3 weeks) (Morris et al., 2015), Shim performed well.

The qualitative data revealed a few subthemes which has been found in other smartphone intervention studies (e.g. Ly et al., 2015), such as the availability and accessibility of the medium. However, most the subthemes were findings which, to our knowledge, have not been found previously. This speaks for the intervention to be something new, novel and distinct from most other type of mental health apps. The most noticeable subtheme with regards to this was the moderating medium, where we categorised quotes indicating the conversational agent made the app work as a moderator for what questions to reflect upon. Two other salient subthemes, which related to limitations of the service were the repetitiveness and shallowness of the app. Interestingly, all these subthemes revealed that many of the participants perceived the app as somewhat a living character, with comments that referred to the app as “a (real) person” and the interaction as a “relationship”, indicating an ability in the chatbot to mimic human guidance. The notion that a bot within the context of health and mental health can develop a bond with its users have been found in previous studies. For example, Bickmore et al. (2005) demonstrated that individuals using a bot to encourage physical activity developed a measurable therapeutic bond with the conversational agent after 30 days. Also, Fitzpatrick et al. (2017) concluded that their chatbot could mirror some therapeutic process, such as empathy and accountability.

4.2. Limitations

A number of limitations need to be highlighted. Firstly, the small sample size necessitates replications of the study with more participants in order to conclude whether the intervention is effective or not with regards to increasing well-being and reducing stress. In addition, the length of the intervention was short due to the lack of content. A new study should let the participants continue using the app for a longer time and even as long as they want. Follow-up measurements after one, three and six months should be included to investigate if the gains are sustained over time. Another limitation was the lack of an active control group to rule out the possibility of placebo effects. Also, studies have shown that it is easier to find significant effects when comparing with a wait-list control rather than a treatment-as-usual control (e.g. Hoermann et al., 2017). However, it has been suggested (Wood et al., 2010) that a wait-list control group can be a highly effective comparison group, especially for this type of intervention since the alternative is often that people do nothing.

The exploratory nature of the study motivated the use of a mixed methods of quantitative and qualitative data. However, the results of

the qualitative data cannot be regarded as representative for all people who receive a smartphone app, built as a conversational agent to deliver a positive psychology and CBT intervention. The sample was limited, small and selected, which means that the findings cannot be generalized. A larger number of participants would have been preferable in order to gain a fuller description. Despite the limitations, the current paper was a first pilot trial to reveal important information on how to continue working with the format in general and the intervention in particular.

5. Conclusions

The current study adds to the body of knowledge regarding both participants' experience as well as the outcome of interacting with a conversational agent for promoting mental health. The results of this pilot study point in a direction that this intervention can be highly engaging and at the same time improve well-being and reduce stress for a non-clinical population. In particular, we believe the good adherence rate, compared to other similar studies and interventions, is promising. Thus, the results validate the usefulness of replicating this study in the future with a larger sample size and an active control group. This is important, as the search for fully automated, yet highly engaging and effective digital self-help interventions for promoting mental health is crucial for the public health.

Funding

This study was sponsored in part by a grant to professor Andersson from Linköping University.

References

- Andersson, G., 2016. Internet-delivered psychological treatments. *Annu. Rev. Clin. Psychol.* 12, 157–179.
- Bergsma, A., 2008. Do self-help books help? *J. Happiness Stud.* 9 (3), 341–360.
- Bickmore, T., Gruber, A., Picard, R., 2005. Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Educ. Couns.* 59 (1), 21–30.
- Bolier, L., Haverman, M., Westerhof, G.J., Riper, H., Smit, F., Bohlmeijer, E., 2013. Positive psychology interventions: a meta-analysis of randomized controlled studies. *BMC Public Health* 13 (1), 119.
- Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. *Qual. Res. Psychol.* 3 (2), 77–101. <http://dx.doi.org/10.1191/1478088706qp0630a>.
- Carlbring, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Öst, L.-G., Andersson, G., 2007. Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Comput. Hum. Behav.* 23 (3), 1421–1434.
- Christensen, H., Griffiths, K.M., Farrer, L., 2009. Adherence in internet interventions for anxiety and depression: systematic review. *J. Med. Internet Res.* 11 (2), e13.
- Cohen, S., Kamarck, T., Mermelstein, R., 1994. Perceived stress scale. In: *Measuring Stress: A Guide for Health and Social Scientists*.
- Dahl, J., Lundgren, T., 2006. Acceptance and commitment therapy (ACT) in the treatment of chronic pain. In: *Mindfulness-Based Treatment Approaches: Clinician's Guide to Evidence Base and Applications*, pp. 285–306.
- den Boer, P.C.A.M., Wiersma, D., van den Bosch, R.J., 2004. Why is self-help neglected in the treatment of emotional disorders? A meta-analysis. *Psychol. Med.* 34 (6), 959–971.
- Diener, E.D., Emmons, R.A., Larsen, R.J., Griffin, S., 1985. The satisfaction with life scale. *J. Pers. Assess.* 49 (1), 71–75.
- Diener, E., Wirtz, D., Biswas-Diener, R., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., 2009. New Measures of Well-Being. *Assessing Well-Being*. pp. 247–266.
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., Biswas-Diener, R., 2010. New well-being measures: short scales to assess flourishing and positive and negative feelings. *Soc. Indic. Res.* 97 (2), 143–156.
- Dunn, E.W., Aknin, L.B., Norton, M.I., 2008. Spending money on others promotes happiness. *Science* 319 (5870), 1687–1688.
- Fava, G.A., Rafanelli, C., Cazzaro, M., Conti, S., Grandi, S., 1998. Well-being therapy. A novel psychotherapeutic approach for residual symptoms of affective disorders. *Psychol. Med.* 28 (2), 475–480.
- Fitzpatrick, K.K., Darcy, A., Vierhile, M., 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment. Health* 4 (2), e19. <http://dx.doi.org/10.2196/mental.7785>.
- Fordyce, M.W., 1977. Development of a program to increase personal happiness. *J. Couns. Psychol.* 24 (6), 511.
- Hayes, N., 2000. *Doing Psychological Research*. Taylor & Francis Group Abingdon.
- Hedman, E., Ljótsson, B., Rück, C., Furmark, T., Carlbring, P., Lindefors, N., Andersson, G., 2010. Internet administration of self-report measures commonly used in research on social anxiety disorder: a psychometric evaluation. *Comput. Hum. Behav.* 26 (4), 736–740.
- Hertzog, M.A., 2008. Considerations in determining sample size for pilot studies. *Res. Nurs. Health* 31 (2), 180–191.
- Hoermann, S., McCabe, K.L., Milne, D.N., Calvo, R.A., 2017. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. *J. Med. Internet Res.* 19 (8), e267.
- Holländare, F., Andersson, G., Engström, I., 2010. A comparison of psychometric properties between internet and paper versions of two depression instruments (BDI-II and MADRS-S) administered to clinic patients. *J. Med. Internet Res.* 12 (5).
- Huppert, F.A., 2009. A new approach to reducing disorder and improving well-being. *Perspect. Psychol. Sci.* 4 (1), 108–111.
- Kelders, S.M., Kok, R.N., Ossebaard, H.C., Van Gemert-Pijnen, J.E.W.C., 2012. Persuasive system design does matter: a systematic review of adherence to web-based interventions. *J. Med. Internet Res.* 14 (6).
- Keyes, C.L.M., Grzywacz, J.G., 2005. Health as a complete state: the added value in work performance and healthcare costs. *J. Occup. Environ. Med.* 47 (5), 523–532.
- Kieser, M., Wassmer, G., 1996. On the use of the upper confidence limit for the variance from a pilot sample for sample size determination. *Biom. J.* 38 (8), 941–949.
- Layous, K., Lyubomirsky, S., 2014. The how, why, what, when, and who of happiness: mechanisms underlying the success of positive activity interventions. In: *Positive Emotion: Integrating the Light Sides and Dark Sides*, pp. 473–495.
- Lewandowski Jr., G.W., 2009. Promoting positive emotions following relationship dissolution through writing. *J. Posit. Psychol.* 4 (1), 21–31.
- Ly, K.H., Trüschel, A., Jarl, L., Magnusson, S., Windahl, T., Johansson, R., Carlbring, P., 2014. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. *BMJ Open* 4 (1), e003440 Retrieved from: <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=prem&NEWS=N&AN=24413342%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3902198&tool=pmcentrez&rendertype=abstract>.
- Ly, K.H., Janni, E., Wrede, R., Sedem, M., Donker, T., Carlbring, P., Andersson, G., 2015. Experiences of a guided smartphone-based behavioral activation therapy for depression: a qualitative study. *Internet Interv.* 2 (1), 60–68.
- Mohr, D.C., Duffe, J., Ho, J., Kwasny, M., Cai, X., Burns, M.N., Begale, M., 2013. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. *PLoS One* 8 (8), e70086.
- Morris, R.R., Schueller, S.M., Picard, R.W., 2015. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. *J. Med. Internet Res.* 17 (3), e72.
- Parks, A.C., Della Porta, M.D., Pierce, R.S., Zilca, R., Lyubomirsky, S., 2012a. Pursuing happiness in everyday life: The characteristics and behaviors of online happiness seekers. *Emotion* 12 (6), 1222–1234. <http://dx.doi.org/10.1037/a0028587>.
- Parks, A.C., Schueller, S.M., Tasimi, A., 2012b. Increasing Happiness in the General Population: Empirically Supported Self-Help?
- Schotanus-Dijkstra, M., Peter, M., Drossaert, C.H.C., Pieterse, M.E., Bolier, L., Walburg, J.A., Bohlmeijer, E.T., 2016. Validation of the flourishing scale in a sample of people with suboptimal levels of mental well-being. *BMC Psychol.* 4 (1), 12.
- Schueller, S.M., 2010. Preferences for positive psychology exercises. *J. Posit. Psychol.* 5 (3), 192–203.
- Schutte, N.S., Searle, T., Meade, S., Dark, N.A., 2012. The effect of meaningfulness and integrative processing in expressive writing on positive and negative affect and life satisfaction. *Cognit. Emot.* 26 (1), 144–152.
- Sheldon, K.M., Boehm, J.K., Lyubomirsky, S., 2012. Variety is the spice of happiness: the hedonic adaptation prevention (HAP) model. In: *Oxford Handbook of Happiness*, pp. 901–914.
- Sin, N.L., Lyubomirsky, S., 2009. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. *J. Clin. Psychol.* 65 (5), 467–487.
- Smyth, J.M., 1998. Written Emotional Expression: Effect Sizes, Outcome Types, and Moderating Variables. *American Psychological Association*.
- Weiss, L.A., Westerhof, G.J., Bohlmeijer, E.T., 2016. Can we increase psychological well-being? The effects of interventions on psychological well-being: a meta-analysis of randomized controlled trials. *PLoS One* 11 (6), e0158092.
- Wood, A.M., Froh, J.J., Geraghty, A.W.A., 2010. Gratitude and well-being: a review and theoretical integration. *Clin. Psychol. Rev.* 30 (7), 890–905.