An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: A Non-Randomized Comparative Trial

Abstract

Background: Recent years have seen an increase in the use of internet-based Cognitive Behavioral Therapy in the area of mental health. Although lower effectiveness and higher dropout rates of unguided than those of guided internet-based Cognitive Behavioral Therapy remain critical issues, not incurring ongoing human costs makes it highly advantageous.

Objective: Paying due consideration to the importance of the therapeutic alliance in psychotherapy, the current research involved an evaluation of the development and efficacy, in terms of mental health, of an application with an embodied conversational agent, enabled for use as an internet-based Cognitive Behavioral Therapy preventative mental health measure.

Methods: Analysis of the data from the 191 participants of the experimental group (average age 38.07 years, \( SD = 10.75 \)) and the 263 participants of the control group (average age 38.05 years, \( SD = 13.45 \)) using a 2-way factorial ANOVA (group \( \times \) time) was performed.

Results: There was a significant main effect \( (P=.02) \) and interaction for time on the variable of positive mental health \( (P=.02) \), and for the treatment group, a significant simple main effect was also found \( (P=.002) \). In addition, there was a significant main effect \( (P=.02) \) and interaction for time on the variable of negative mental health \( (P=.005) \), and for the treatment group, a significant simple main effect was also found \( (P=.001) \).

Conclusions: This research can be seen to represent a certain level of evidence for the mental health application developed herein, indicating empirically that the embodied conversational agent can be useful in mental health care. Meanwhile, given the issues in the non-randomized nature of the comparison study, it is necessary to pursue higher quality evidence while continuing to further improve the application, based on the findings of the current research.

Keywords: embodied conversational agent; cognitive behavioral therapy; psychological distress; mental well-being; artificial intelligence technology

Introduction

Current state of mental health care utilizing information and communication technologies, and its significance

In the current era, the use of biological, psychological, and social models [1] to understand and support mental health issues is common, and of these, Cognitive Behavioral Therapy (hereafter CBT) is one of the major means of providing psychological understanding and support. Owing to its confirmed efficacy across a range of mental health care-related fields, from changing daily habits to specialist interventions for mental illness, CBT has become a psychological approach used worldwide and has been delivered in various forms to date, including face-to-face,
group therapy, and books (cf. Westbrook, Kennerley, & Kirk [2]). Meanwhile, there has long been a problem with patients not visiting a clinician or not discussing their concerns, even when there is a mental health issue, and research into the service gap, that is, the difference between the need for and uptake of mental health services, has been ongoing [3]. Underlying this is the dual problem of support providers and practitioners being unable to reach those in need of services, as well as that of service users being unable to access services. The former involves issues of privacy and cost concerning service provision, while the latter deals with the stigma surrounding psychiatry and mental illness, as well as preferences regarding methods of accessing support.

In recent years, internet-based CBT (hereafter iCBT) and computerized CBT have increasingly been used as a means to fill the service gap and resolve various problems related to mental health, such as bipolar disorder [4,5], anxiety disorder [6], depression [7], treatment adherence [8], and common mental health problems [9]. According to Andersson’s review [10,11], therapist-guided iCBT is standard, and while its efficacy in the three characteristic areas of depression, anxiety, and physical symptoms has been shown to be almost equivalent to that of face-to-face CBT, the issues of lower efficacy and higher dropout rates of unguided iCBT than those of guided iCBT have been raised. Even in a systematic review of depression-related self-help smart-phone applications conducted by Huguet et al. [12], it became apparent that there are no suitable, evidenced-based CBT and behavioral action (BA) applications, that is, no unguided iCBT applications are available, despite the large societal demand for and number of applications available.

Potential of mental health care with embodied conversational agent

However, given that service provision cost was one of the factors motivating the original use of iCBT, unguided iCBT is a very attractive option because ongoing human costs are unnecessary. When considered in terms of underlying clinical issues, the therapeutic alliance issue could conceivably underlie the problem of iCBT’s low efficacy and high dropout rate. The therapeutic alliance has been considered an integral element in not just face-to-face CBT, but in all forms of psychotherapy, in terms of its role as a common factor in the efficacy of treatment [13]. The therapeutic alliance, as formed in unguided iCBT, is noted to be lacking in terms of elements of “development” and “maintenance” [14]. As such, artificial intelligence (AI) technology, especially Embodied Conversational Agent (ECA), is conceivable as an effective means of using technology to overcome the paucity of therapeutic alliance-developing elements in unguided iCBT. In fact, its efficacy in reducing symptoms of depression has been reported by Fitzpatrick, Darcy, and Vierhile [15] in their research, utilizing a completely automated text-based response agent based on CBT principles. By offering pseudo-dialog experiences of freestyle dialog with the agent, the therapist-role agent in particular allows for the development of the therapeutic alliance even in unguided iCBT, and thus could contribute to increased efficacy and reduced dropout rates. However, the use of agent technology in the area of mental health care is in its infancy. Historically, virtual affective agents were utilized in serious games for health care [16]. However,
most of these are limited to narrow contexts and do not enable complex natural-language interactions [16]. Thus, currently there is insufficient information regarding the combined use of an ECA that is capable of text-based interactions with traditional iCBT and its efficacy.

Against this background, the current research involved an evaluation of the development and efficacy, in terms of mental health, of an application with an ECA that is capable of text-based freestyle dialog enabled for use in iCBT preventative mental health. In doing so, the investigation was conducted using a non-randomized comparison study, in light of research reporting the general ineffectiveness and high dropout rate of unguided iCBT [10,11], as well as the current general lack of awareness regarding agent use in iCBT. While non-randomized prospective studies (NPSs), such as this non-randomized comparison study, are inferior to randomized comparison studies in terms of quality of evidence, they offer benefits such as significant low risk of discontinuation [17] and also allow the most economic use of resources. Moreover, in accordance with the aims of preventative mental health measures, positive and negative mental health effects indices were used as general mental health indicators.

Methods

Unguided iCBT method

We use the “SABORI” as an iCBT application, which is a self-care application developed by the Laboratory of third author (The University), based on CBT and BA principles, for the purpose of preventing mental health issues. It is a web-based unguided iCBT application available for use on a smart-phone, tablet, or computer browsers. SABORI users engage in self-monitoring by answering questions about changes in their daily mood and physical condition, subsequently receiving feedback and behavioral suggestions relevant to their responses. All user interactions in answering questions were only for the user to choose options; text-based interaction was not available. Improving one’s own mental health state thus becomes intentional, through the promotion of self-monitoring and behavioral activation [18]. In this work, we improved and upgraded the interactive contents on the SABORI application to achieve better adherence. The two major reasons are as follows. First, the structure of SABORI was suited to adding a conversation-style dialog capability, because of the format of a user answering questions of the agent in the application. Second, since results from the preliminary study without a control group suggested improvement in mood due to behavioral activation for the depression group from one month of use [18], it was considered that addition of a freestyle dialog capability would be unlikely to adversely affect users. In this study, we upgraded SABORI by adding an agent-based freestyle dialog capability to the behavior suggestion section of SABORI, and investigated its psychological effects. Figures S1 and S2 show screenshot examples of freestyle dialog during monitoring and behavioral suggestions.
Development of the conversation-style dialog capability on iCBT application

First, the authors sought and gained ethics approval from our University Ethics Review Committee for the entire research plan, including the study. The majority of the development and implementation was conducted by two university lecturers specializing in clinical psychology, one information engineering specialist, and two specialists from a dialog system development company. The dialog system used in this research is an AI technology system, which includes Multi-Agent System as an AI engine, constructed from numerous agents with individual dictionaries or rules defined for each domain. The basic structure of conversation is that first the agent asks the user a behavior suggestion-related question and then responds to the user’s input, and following this sequence, it transitions naturally to SABORI’s existing behavior suggestion (Multimedia Appendix 1,2). The ontology for the system to classify user input and the creation of the behavior-related questions were constructed from September to December 2016. Basic knowledge for conversation was also collected from internet, for example Wikipedia.

Human-in-the-loop improvement of the freestyle dialog feature

In March 2017, a preliminary study was conducted to test and improve the freestyle dialog capability. The freestyle dialog capability was revised to form natural conversation patterns by having 10 clinical psychology students use it for one-month, allowing it to learn vocabulary and phrases not initially predicted and by altering the system’s behavior suggestion method. In addition, during this period, 1162 cases of user input were gathered, and the system’s response success rate was 66.52% at this stage. Moreover, the response success rate was determined as the total number of input cases excluding those that it was unable to respond to.

Furthermore, in the preliminary study, when the behavioral suggestion was provided using the user’s unadulterated input, the uniqueness of the pattern was conspicuous to continuing users, and while the system was aware of this, from the user’s perspective, it appeared that no change had occurred, thus creating the impression in many users that the system was unaware. Therefore, in consideration of the differences between the system’s awareness rates and the user’s perceived system awareness rates, official names were used where possible, and responses were changed to emphasize awareness of natural conversation.

Procedure and Participants

In June 2017, the services of an internet research company were engaged, for the recruitment of predicted iCBT application users, while being mindful to find almost equal numbers of company employees, university students, and housewives. The selection process is outlined in Figure 1.
A total of 10,963 individuals were sent recruitment notices explaining, “SABORI is a self-monitoring service based on CBT, which is recognized for its high efficacy in mental health care,” along with materials outlining application and were asked to choose between “I would like to use it” and “I would not like to use it.” Of these, 2,668 individuals remained after removing the 8,295 individuals who did not wish to use the application. Following this, explanatory sessions were offered over two days to explain the rules of use during the current month-long research study, namely the condition of using the application more than once every two days, totaling over 15 days of use. After removing 2,109 individuals who were unable to participate, the remaining 559 individuals formed the experimental sample for the current research. After the one-month usage period, of the 427 responses from the post-study questionnaire, it was revealed that 191 participants in the experimental group complied with the rule stipulating more than 15 days of use (46 male office workers, 34 female office workers, 13 male university students, 13 female university students, and 85 housewives; average age 38.04 years, $SD = 10.75$). A total of 236 individuals used the application for less than 15 days. In addition, the 2,109 individuals who were unable to participate in the one-month usage period were designated as the control group, and responses were closed when almost equivalent
numbers of post-study questionnaire responses were received, forming a control
group of 263 individuals (51 male office workers, 53 female office workers, 26 male
university students, 40 female university students, and 93 housewives; average age
38.05 years, SD = 13.45). Moreover, members of the treatment group were paid the
equivalent of $4 USD worth of points from the internet research company, for their
participation. As a result, in response to the 6067 cases of user input accumulated
during the study, the system’s response success rate significantly improved at
92.93%.

Measures

1. World Health Organization-Five Well-Being Index (WHO-5-J)
The Japanese version of the World Health Organization-Five Well-Being Index
(WHO-5-J), developed by the World Health Organization (WHO), released by
Psychiatric Research Unit, Psychiatric Center North Zealand [19], and translated by
Awata et al. [20], was utilized in this study. It is a one-factor scale measuring positive
mental health related to physical aspects using a five-item, six-point Likert scale. The
reliability and validity of the Japanese version has been confirmed by Awata et al.
[20]. It was used as an evaluative index utilizing total scores, referring to the past
two weeks.

2. Kessler 10 (K10), Japanese Version
A one-factor scale measuring negative mental health related to physical aspects
using ratings of 10 items on a 5-point Likert scale was created by Kessler et al. [21].
The reliability and validity of the Japanese version has been confirmed by Furukawa,
Ohno, Uda, and Nakane [22]. It was used as an effect index utilizing total scores
referring to the past two weeks.

3. Behavioral Activation for Depression Scale (BADS), Japanese Version
A scale by Kanter, Mulick, Busch, Berlin, and Martell [23] was used, measuring four
Impairment” (BADS-WS), and “Social Impairment” (BADS-SI), over 25 items rated on
a 7-point Likert scale, designed to evaluate functional impairment, avoidance, and
activation in behavioral activation. The reliability and validity of its Japanese version
has been confirmed by Takagaki et al. [24]. It was used as an effect index of
activation (BADS-AC) over 7 items, and avoidance/rumination (BADS-AR) over 8
items, utilizing total scores for each subscale, referring to the past two weeks. We
improved SABORI application to support behavioral activation based on daily mood
and physical condition and hypothesized that conversation-style free dialog
promotes behavioral activation and reduces avoidance.

Furthermore, a total of 48 scale items unrelated to the aims of the study, including
items not mentioned here, were removed.
Results

Effect index descriptive statistics

Table 1 details descriptive statistics for both groups’ pre- and post-treatment scores.

Table 1. Mean, Standard Deviation, and Cronbach’s α of Outcome Measures

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th></th>
<th></th>
<th>Intervention</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>α</td>
<td>Mean</td>
<td>SD</td>
<td>α</td>
</tr>
<tr>
<td>WHO-5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-test</td>
<td>15.64</td>
<td>5.53</td>
<td>.91</td>
<td>15.03</td>
<td>5.26</td>
<td>.92</td>
</tr>
<tr>
<td>post-test</td>
<td>15.65</td>
<td>5.35</td>
<td>.91</td>
<td>16.12</td>
<td>4.71</td>
<td>.89</td>
</tr>
<tr>
<td>K10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-test</td>
<td>23.76</td>
<td>9.97</td>
<td>.95</td>
<td>23.58</td>
<td>9.56</td>
<td>.94</td>
</tr>
<tr>
<td>post-test</td>
<td>23.97</td>
<td>9.89</td>
<td>.95</td>
<td>21.56</td>
<td>8.26</td>
<td>.93</td>
</tr>
<tr>
<td>BADS-AC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-test</td>
<td>15.67</td>
<td>8.27</td>
<td>.89</td>
<td>16.09</td>
<td>8.36</td>
<td>.88</td>
</tr>
<tr>
<td>post-test</td>
<td>15.84</td>
<td>8.66</td>
<td>.84</td>
<td>17.19</td>
<td>7.90</td>
<td>.84</td>
</tr>
<tr>
<td>BADS-AR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-test</td>
<td>17.71</td>
<td>8.75</td>
<td>.91</td>
<td>18.51</td>
<td>8.79</td>
<td>.89</td>
</tr>
<tr>
<td>post-test</td>
<td>18.35</td>
<td>9.36</td>
<td>.88</td>
<td>17.84</td>
<td>8.03</td>
<td>.82</td>
</tr>
</tbody>
</table>

WHO-5: WHO-Five Well-Being Index
K10: Kessler 10
BADS-AC: Behavioral Activation for Depression Scale - Activation
BADS-AR: Behavioral Activation for Depression Scale - Avoidance/Rumination

Results of the 2-factor ANOVA (Group × Time)

A two-factor mixed model ANOVA was conducted to investigate differences for each effect index, group (treatment/control), and time (pre/post) according to each scale (Table 2, Figure 2-5). Moreover, error bars indicate a 95% confidence interval for the mean.

Table 2. Effect of Interaction between Time and Group on Outcome Variables
<table>
<thead>
<tr>
<th></th>
<th>Main Effect</th>
<th>Interaction Effect</th>
<th>simple main effect (partial $\eta^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time</td>
<td>Group</td>
<td>Time x Group</td>
</tr>
<tr>
<td>WHO-5</td>
<td>5.79</td>
<td>.02</td>
<td>5.47</td>
</tr>
<tr>
<td>K10</td>
<td>5.43</td>
<td>2.52</td>
<td>8.11</td>
</tr>
<tr>
<td>BADS-AC</td>
<td>2.75</td>
<td>1.45</td>
<td>1.66</td>
</tr>
<tr>
<td>BADS-AR</td>
<td>.00</td>
<td>.40</td>
<td>2.39</td>
</tr>
</tbody>
</table>

WHO-5 : WHO-Five Well-Being Index
K10 : Kessler 10
BADS-AC : Behavioral Activation for Depression Scale - Activation
BADS-AR : Behavioral Activation for Depression Scale - Avoidance/Rumination

**Fig2.** Change in WHO-5
Fig3. Change in K10

Fig4. Change in BADS-AC
WHO-5 results indicated a significant main effect and interaction for time, $F(1,452) = 5.79, P=.02$; $F(1,452) = 5.47, P=.02$, respectively. After investigating the simple main effect of time for each level of the groups, a significant simple main effect was found for the treatment group, $F(1,452) = 9.71, P=.002$. On the K10, there was a significant main and interaction effects for time, $F(1,452) = 5.43, P=.02$; $F(1,452) = 8.11, P=.005$, respectively. Again, after investigating the simple main effect of time for each level of the groups, a significant simple main effect was found for the treatment group, $F(1,452) = 11.57, P=.001$. The BADS-AC results indicated that there was a significant main effect for time, $F(1,452) = 2.75, P=.098$. Investigation of the simple main effect of time for group each level indicated a significant simple main effect trend in the treatment group, $F(1,452) = 3.53, P=.06$. Neither a significant main nor interaction effects were found on the BADS-AR.

**Discussion**

**Positive and negative effects on mental health**

Results of the two-way ANOVA (group and time independent variables) indicated that the SABORI had an effect on the two variables: positive and negative mental health. Accordingly, this can be seen as evidence suggesting the efficacy of the agent-based conversation-style mental health self-care application, SABORI, across a wide range of areas, due to the indicated effects on both positive and negative mental health. Furthermore, research into the efficacy of the version of SABORI without the conversation-style dialog feature [18] found that it improved negative mental health for depression only, whereas the current research indicated that the new version of SABORI application with an ECA is effective beyond just depression, acting across a
wide range of areas, including preventive mental health measures. The addition of
the agent-based dialog feature potentially affected the strengthening of the
therapeutic alliance between the system and the user, and as such, there is a need for
further detailed research into the factors underlying the effect, as well as the
component factors of the therapeutic alliance when utilizing agent. Agent, especially
the ECA, is a technology that has the potential to vastly change traditional mental
health support measures, through its use as an alternative to face-to-face interviews,
and the current research findings are of major significance to future development,
given that the research has confirmed the efficacy of adding ECA to unguided iCBT,
which has traditionally been criticized on grounds of low effectiveness and paucity
of evidence. Nevertheless, in light of the still inadequate efficacy level, we must
continue to seek ways to increase efficacy by linking to educational psychology
content and delivering optimal individualization of the dialog-based behavior
suggestion feature, along with further improvement of knowledge-base of ECA with
the data collected in this study. Moreover, given that the current research only
involved a non-randomized comparison study, inferiority in terms of quality of
evidence is undeniable with regard to the issue of evidence from randomized
comparison research. Consequently, based on the findings of the present research,
there is a need to pursue a higher level of evidence, by conducting a randomized
comparison study and adding different participant groups.

Efficacy regarding behavioral activation, avoidance, and rumination

The two-way ANOVA (independent variables: group × time) found a significant
trend for behavioral activation, suggesting the potential for a certain degree of
effectiveness. One possible reason for the weakness of the behavioral activation
effect was the one-month usage time frame of the current research. Although the
new version of SABORI offers behavioral suggestions through the dialog feature
based on current circumstances, these function as a catalyst at most, and the intent
is that, with continued usage, the user will become able to undertake behavioral
activation. However, it is conceivable that the one-month time frame is too short for
voluntary behavioral activation to arise, so the effect never extends further than
carrying out the behavioral suggestions. Future research will need to investigate
how behavioral activation changes over a longer period of time, based on monthly
use. In addition, no effect was found for behavioral activation related factors of
avoidance and rumination. There are two possible reasons for this. First, similar to
behavioral activation, the one-month time frame may have been too short to elicit a
behavioral change, revealing the necessity for a further longer-term study. Second, it
is conceivable that the use of healthy participants may have contributed to this
outcome. Since the newly developed SABORI is intended for use as a mental health
preventative measure, it has no particular audience limitations. Therefore, it is
possible that any effect could not be adequately measured in a healthy cohort, given
that behavioral activation treatment is primarily intended as a treatment for
depression, and the BADS used in the current research as an effect index also
contains many items relating to depression. Consequently, it is hoped that the
efficacy of the iCBT application with an ECA will be even more comprehensively
examined by conducting research utilizing effect indices that are more suitable in terms of mental health preventative measures intended for a healthy cohort.

Conclusions

The current research provides evidence regarding mental health care delivered by the SABORI, a self-care application utilizing AI technology, especially agent technology. To date, unguided iCBT has failed to garner attention, being considered ineffective with high dropout rates [10,11]. However, current research has revealed that it is possible to use ECA to compensate for clinical failures of iCBT such as the impoverished therapeutic alliance. If the technological development of conversational agent continues to progress, it could be possible to form a therapeutic alliance with agent to rival that found in face-to-face therapy. The utilization of agent technology is anticipated to vastly change the traditional, mainstream delivery of mental health services and could be an innovation that holds the secret to filling the service gap [3]. Moreover, it is expected to contribute greatly to clinical practice and be increasingly utilized in the area of mental health care.

Acknowledgements

We would like to thank Mr. Yuta Tonoike and Mr. Daiki Morita of Mind Isle Co. Ltd for their cooperation in this research study. This research was supported by Graduate Program for Social ICT Global Creative Leaders (GCL) of The University of Tokyo by Japan Society for the Promotion of Science.

Conflicts of Interest

None declared.

Abbreviations

AI: Artificial Intelligence
ANOVA: Analysis of variance
BADS: Behavioral Activation for Depression Scale
BADS-AC: Behavioral Activation for Depression Scale - Activation
BADS-AR: Behavioral Activation for Depression Scale - Avoidance/Rumination
BADS-WS: Behavioral Activation for Depression Scale - Work/School Impairment
BADS-SI: Behavioral Activation for Depression Scale - Social Impairment
CBT: Cognitive behavioral therapy
ECA: Embodied Conversational Agent
iCBT: internet-based CBT
K10: Kessler 10
NPS: Non-randomized Prospective Study
WHO-5-J: World Health Organization-Five Well-Being Index
Multimedia Appendix 1 Screenshot of the monitoring questions
(Involving 16 items concerning one’s own mood and physical condition. The question “How did you feel when you woke up this morning?” can be answered by choosing “Great,” “Fine,” or “Not good.”)

Multimedia Appendix 2 Screenshot of the behavioral suggestion dialog.
(“Activation,” “physical activation,” “psychological activation,” or “suppressed behavior” proposals will be chosen depending on the monitoring results, and the user and AI will then engage in related dialog. This screenshot is an example of “suppression.” “Seems like you don’t have much energy. Maybe you should go to bed a little earlier than usual. Even simply lying down can make a big difference. Did you sleep well last night?” If you respond with “I slept well last night,” AI will counter this with, “That’s good. Let’s try hard again today!” The behavioral suggestion, “Today’s goal: Maybe you should go to bed a little earlier than usual. Even simply lying down can make a big difference.” is then displayed below).

References
of Psychopathology and Behavioral Assessment. 2007;29: 191–202. PMID:11126230