

# Exploring the Effects of Incorporating Human Experts to Deliver Journaling Guidance through a Chatbot

YI-CHIEH LEE, University of Illinois at Urbana-Champaign & NTT Communication Science Laboratories  
 NAOMI YAMASHITA, NTT Communication Science Laboratories, Japan  
 YUN HUANG, University of Illinois at Urbana-Champaign, USA

Chatbots are regarded as a promising technology for delivering guidance. Prior studies show that chatbots have the potential of coaching users to learn different skills; however, several limitations of chatbot-based approaches remain. People may become disengaged from using chatbot-guided systems and fail to follow the guidance for complex tasks. In this paper, we design chatbots with (HC) and without (OC) human support to deliver guidance for people to practice journaling skills. We conducted a mixed-method study with 35 participants to investigate their actual interaction, perceived interaction, and the effects of interacting with the two chatbots. The participants were randomly assigned to use one of the chatbots for four weeks. Our results show that the HC participants followed the guidance more faithfully during journaling practices and perceived a significantly higher level of engagement and trust with the chatbot system than the OC participants. However, after finishing the journaling-skill training session, the OC participants were more willing to keep using the learned skills than the HC participants. Our work provides new insights into the design of integrating human support into chatbot-based interventions for delivering guidance.

CCS Concepts: • **Applied computing** → *Psychology*; • **Human-centered computing** → **User studies**.

Additional Key Words and Phrases: Chatbot; Conversational Agent; Self-disclosure; Human Support; Training

## ACM Reference Format:

Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2021. Exploring the Effects of Incorporating Human Experts to Deliver Journaling Guidance through a Chatbot. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 122 (April 2021), 27 pages. <https://doi.org/10.1145/3449196>

## 1 INTRODUCTION

Chatbots are regarded as one of the most promising technologies and are increasingly applied in many domains. Because chatbots provide a fast, convenient, and low-cost communication channel, both scholars and practitioners are keen to develop effective chatbots to address the challenges of providing healthcare services. For example, a growing body of research demonstrates how chatbots can be useful for helping people maintain good lifestyles [29, 37], collecting daily health information to share with healthcare providers [26, 34], and guiding people to improve their general well-being [28, 32, 63]. For instance, Wang et al. [61] proposed a conversational agent to coach people to relieve their public-speaking anxiety through cognitive reconstruction exercises, and Fitzpatrick et

Authors' addresses: Yi-Chieh Lee, University of Illinois at Urbana-Champaign & NTT Communication Science Laboratories, ylee267@illinois.edu; Naomi Yamashita, NTT Communication Science Laboratories, Japan, naomiy@acm.org; Yun Huang, University of Illinois at Urbana-Champaign, USA, yunhuang@illinois.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2021 Association for Computing Machinery.

2573-0142/2021/4-ART122 \$15.00

<https://doi.org/10.1145/3449196>



al.'s [17] Woebot system gives step-by-step guidance for users to think through their situation with cognitive behavioral therapy and was found to relieve users' depression. Other recent studies have applied a variety of conversational strategies and structures to promote behavioral change and to persuade chatbot users to act differently [27, 29, 57]. Some of these systems have even been found to outperform human-human interaction in some scenarios. For example, Lucas et al. [36] found that utilizing a virtual agent as an interviewer could promote users' depth of self-disclosure, and Xu et al. [67] concluded that the use of interactive robot agents would probably enhance physical-therapy outcomes. Therefore, these prior works have demonstrated that chatbots can serve as an effective platform for delivering guidance and tutoring people.

Despite the success of utilizing chatbots to deliver guidance, there are still a number of challenges to overcome. For example, research points out that people easily become disengaged from using a chatbot [48, 58], hampering them from long-term interventions. Moreover, people may overtrust solutions suggested by chatbots which could be inappropriate [24, 67, 68]. In another study, Luria et al. [38] found that people felt uncomfortable interacting with a chatbot which used the same personality to handle both low-risk (e.g., social chat) and high-risk (e.g., medical purpose) contexts. Thus, the authors suggested to design a chatbot that embodies multiple personalities, each of which are displayed in a unique social presence and have the expertise to focus on a single task.

Prior studies inspire us to overcome challenges by integrating human support into a chatbot system. More specifically, we may be able to make the best use of both human-based and chatbot-based approaches by co-embodying them into a single system. Indeed, studies have suggested that the integration of human support with chatbot interactions could promote user engagement [48] and efficacy of using self-guided systems. For example, a recent study [34] proposed a mediator chatbot that promotes deep self-disclosure from users and delivers the information to a human expert. More research is clearly needed on how individuals might respond differently to interaction with a chatbot alone vs. one incorporating human support. We are also interested in understanding how such differences affect user experience in the long run. To help fill the gap, we conducted a mixed-methods study with 35 participants. We deployed two chatbot designs, both of which delivered training in journaling skills [22, 25]. The first version of the chatbot guided participants in the journaling skills itself, while the second version integrated a human expert (coach) into its interaction when guiding the participants in the journaling skills. Over a period of four weeks, we tracked changes and differences in how each version impacted users' responses to and perceptions of the chatbot system, as well as their level of compliance with the guidance to practice journaling skills.

Our work makes several contributions to the CSCW and human-computer interaction (HCI) communities. It is among the first that investigated the effects of integrating a human expert to deliver guidance for practicing journaling skills. Our unique three-phase design of an experimental study with 35 participants contributes novel findings of how chatbot interactions with and without expert guidance elicited user interaction differently over time. More specifically, during the *Training* phase, participants' actual and perceived engagement with the chatbot providing expert guidance (HC) was significantly higher than that of the participants who interacted with their chatbot alone (OC); however, during the *Free-will* phase, the OC participants chose to continue practicing journaling significantly more than the HC participants. Second, triangulating system log analysis with interviews and surveys, we provide new insights into how the design of chatbot systems with and without human support affected user experience of such systems both objectively (e.g., in terms of the length and depth of journaling content) and subjectively (e.g., participants' perceived trust and intimacy with the chatbots). Third, our work also presented empirical evidence of using chatbots to practice journaling on improving participants' self-reflection. Since prior work shows better self-reflection could enhance people's awareness of their well-being, our work further provides



design implications for applying chatbots in the healthcare domain and to support diverse training purposes.

## 2 RELATED WORK

### 2.1 Chatbot for Delivering Guidance

Conversational agents (e.g., chatbots) are gaining considerable attention in many fields including healthcare [69] and education [64]. Research has shown that chatbots can assist users in tracking and monitoring their behavior (e.g., [37]) and feelings (e.g., [17]), which could further be used to solicit social support and self-reflection [29]. Also, many studies designed chatbots to guide healthier habits or ways of thinking [46], such as better eating habits [37], exercise [29], ways of coping with stress [46], and self-compassion [32]. For example, Park et al. [46] incorporated a motivational interview technique into chatbot conversation to help users cope with stress, and found that their design facilitated conversations that improved self-reflection as well as stress management. Lee et al. [32] designed a dialogue aimed at inspiring users to take care of a chatbot that was portrayed as having had a negative experience, and found that after doing this for two weeks, users' self-compassion increased significantly. Another line of research has shown that chatbots have the potential to help people improve their mental well-being by training their thoughts and behavior [36, 61, 65]. For instance, Wang et al [61] designed a public-speaking tutor using a chatbot system to coach users and reduce their public speaking anxiety. Hence, these studies have shown that chatbots could not only help track users' behavior but could also play a proactive role in training users to learn skills.

Recent advancements in artificial intelligence (AI) are also enabling chatbots and other virtual agents to act more credibly like human beings, including during the provision of self-help information [10, 14]. Prior studies [5, 33] indicated that conversational interaction can increase trust and affect users' acceptance of recommendations from a conversational agent. Thus, the design of the interaction between them is important in enhancing users' willingness to adopt chatbot suggestions. Gabrielli et al. [18] proposed a chatbot-based coaching intervention that successfully helped adolescents learn life skills, such as strategies for coping with bullying, and previous research [24, 67] found that their participants' trust and compliance with physical therapeutic suggestions were both higher when interacting with robot therapy partners than with a human expert. Moreover, research has shown that people tend to apply the social norms of human relationships to their interactions with computer agents. This tendency, known as the Computers Are Social Actors (CASA) paradigm [42], has informed the design of many computer agents [35, 50, 57]. People may perceive intimacy and companionship with a computer agent [27, 35, 40], inducing changes in behavior change. For example, Ravichander et al. [50] found that reciprocity occurred in human-chatbot interactions and that a chatbot's self-disclosure encouraged people's self-disclosure. Similarly, recent work by Lee et al. [35] showed that a chatbot's self-disclosure improved participants' perceived intimacy with the chatbot and facilitated their self-disclosures in response to the chatbot's sensitive questions (e.g., failure experiences).

However, several limitations of chatbot-based approaches remain, and in certain situations, chatbot-based approaches may be less beneficial than those provided by humans [40, 41, 48]. For example, Howard et al. [24] has pointed out that some people may trust robots too much, due to over-optimism about the viability of the solutions they suggest, and that this trust becomes a source of risk if robots make clinically suboptimal or inappropriate suggestions. In addition, for healthcare interventions that require long-term engagement [8], people may easily become disengaged from the use of self-guided systems, due to loss of motivation and/or failure to incorporate those systems' recommendations into their daily lives [48]. Furthermore, an investment model shows



that purely computer-based interventions are often much less effective than hybrid ones with some professional human input [16], in part because the latter tends to inspire their users to execute a higher proportion of their intervention requests.

## 2.2 Integration of Human (Expert) Support and Chatbot-based Approaches

In prior works, human support has been provided via a separate communication channel external to the chatbot system, such as phone calls, text messages, and email [1]. For example, there have been two main ways of providing human support to chatbot systems. The first is to deploy chatbots in between human-run sessions, to offer users unbroken access to materials and activities [52, 53, 69]. Studies that have adopted such an approach regard chatbots as supplementary tools to support human expert's intervention; chiefly, by monitoring clients outside of their clinical sessions, and garnering information about them that may result in better treatment (e.g., [28, 48]). Alternatively, it is possible to design a primarily chatbot-based intervention, augmented by human supporters who promote engagement and provide technical troubleshooting and clinical support when issues arise [41, 48, 60]. Such an approach could be more efficacious than interventions by chatbots unsupported by humans [48, 52].

Recently, some researchers have suggested an integration of human support into chatbot interventions [48, 52, 53]. For example, Schueller et al. [52, 53] reviewed prior studies of integrating human experts (e.g., coach and therapist) into behavioral intervention technologies, not chatbot-based, and suggested concepts to guide a deeper integration by capturing the trade-offs between client benefits and the available human resources. Alternatively, some prior studies [15, 31] use conversational agents to encourage users' collaboration and communication between people. Specifically, Kumar et al. [31] designed a chatbot tutoring system which gave guidance for multiple students to facilitate collaborative learning among them. Duan et al. [15] utilized a conversational agent to enhance non-native speakers' confidence in conversation with native speakers. These studies showed that chatbots could help mediate interactions between users, but we further explore the effect of applying chatbots to mediate suggestions for guiding users to learn skills.

The foregoing review and Schueller et al.'s work [52, 53] calls for an integration of the support provided by chatbots and humans. However, previous works have indicated different designs for implementing multiple personas into a chatbot system. For example, Luria et al. [38] conducted studies to examine multiple personalities for conversational agents under different contexts (e.g., low-risk and high-risk contexts). They found that users preferred to have an additional expert agent guiding a specific complex task instead of interacting with the same agent that handled both simple and complex tasks. Conversely, Chaves et al. [11] found that users reported confusion when they engaged in multiple persona chatbots for an information gathering task in a single communication channel. Therefore, it is not clear if and how adding a human supporter into human-chatbot interactions could impact user experience and outcomes, which motivates our research.

## 3 RESEARCH QUESTIONS

To explore the effects of integrating human support into a chatbot system to deliver guidance, we examine two chatbot designs, one with and one without integrating human support. The chatbots are designed to guide users to learn journaling skills. Journaling is an approach suggested to help improve mental health [22, 25]. We chose gratitude journals [13, 55] and expressive writing [3] as the journaling skills because they are known to be effective in improving self-reflection and mental well-being. Users' journaling exercises could be used to measure their compliance with a chatbot's guidance. We conducted a four-week study deploying two chatbot conditions, with (HC condition) and without (OC condition) a human supporter (coach), to provide guidance to learn the



journaling skills. The suggestions delivered to the users in both chatbot conditions were adopted by pre-existing journaling materials (e.g., gratitude journals [13, 55] and expressive writing [3]). The evaluation of how each design influenced its users' experience and journaling behavior was guided by the following research questions.

**RQ1:** *Do people interact with their chatbot differently if they have a human expert (HC) or not (OC) to guide them?* The goal of our design is to understand whether integrating a human expert in the loop of interaction may affect users' journaling practices by following the suggestions delivered through the chatbot system. Previous research [36] suggests that people disclose more deeply to chatbots than to a human interviewer, but research [38, 48] found that people might feel uncomfortable following suggestions for complex tasks when it is given by a chatbot. Thus, we evaluate users' depth of disclosure to measure how users would follow guidance for journaling. Our research findings of RQ1 will contribute design insights of incorporating human experts into human-chatbot interaction for delivering guidance.

**RQ2:** *How do people perceive their interaction with the chatbot differently between the HC and OC conditions?* To address RQ2, we studied two perspectives: **a)** people's perceived interaction with the chatbots in general, and **b)** people's perceived benefits of practicing journaling through the chatbots. First, when examining their perceived interaction with the chatbot system, we applied several constructs, including people's perceived engagement, trust, and intimacy. The reasons we measured these constructs are as follows. Prior research [57] showed that when users identified a chatbot as a human, they would think that the conversation was more engaging and persuasive. Therefore, we expect that incorporating a human expert in delivering guidance for learning journaling skills would improve people's perceived engagement. Moreover, previous studies found that users' trust [5, 33] in and intimacy [27, 35] with the conversational agents would affect their behavior to accept the suggestions and disclose themselves. Thus, to understand different perspectives which may influence users' behaviors of practicing the journaling skills, we measure their perceived engagement, trust, and intimacy before and after being given guidance for practicing journaling skills. Second, practicing journaling skills [22, 25] may help improve self-reflection and higher levels of self-awareness [25, 54, 59], which leads to improved behavioral changes. For example, prior research [29] showed that a chatbot could guide users to better self-reflect on their physical activities. Therefore, we also measured users' perceived self-reflection and self-awareness to explore whether the two chatbot designs to deliver journaling suggestions would affect users' perceptions differently.

**RQ3:** *Do people keep practicing journaling skills differently over time between the HC and OC conditions?* Finally, we further explore how the designs would influence users' willingness to keep practicing those suggestions as a measure of the lasting effect of the design. Prior research [35, 38] suggests that when people interact with an agent over an extended period of time, their familiarity with the system may affect further behaviors. In addition, prior research [8, 69] indicated that people may easily become disengaged from the use of self-guided systems. Thus, we explore how our designs, with and without human support, affect users' retention of their journaling exercises over time.

## 4 METHOD

### 4.1 Study Design

In this study, each participant was asked to interact with the chatbot every day, via a single daily chatting task that lasted approximately 15-30 minutes for four weeks. The goal of the chatbot system was guiding participants to practice journaling skills.



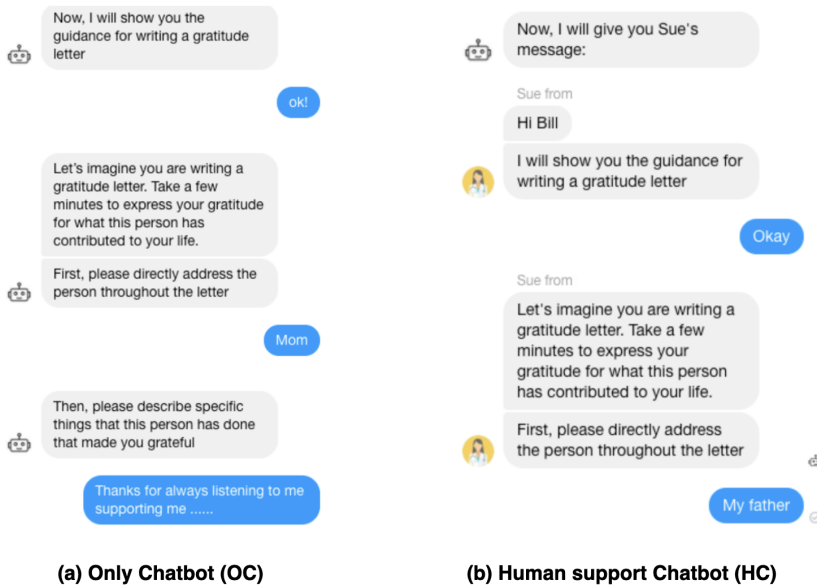


Fig. 1. Our chatting interface. (a) The chatbot gives a suggestion/guidance to the user (OC group); (b) The chatbot introduces a human coach, and the coach's agent gives a suggestion/guidance to the user (HC group). The chatbot switched the icon and name to the coach only when the users were in the *Suggestion* session (Fig. 2).

Two experimental conditions were designed: an Only Chatbot (OC) condition in which the participants interacted with a chatbot for the four-week study, and a Human support (coach) with Chatbot (HC) condition in which the participants interacted with a chatbot for the four-week study, but the chatbot would introduce a human coach when delivering suggestions for journaling skills (Fig. 1). The participants in the HC condition were informed that their journaling suggestions were given by a human expert (coach). However, they were told that the coach was too busy to chat with them in real-time, so the coach would leave the suggestions at night and the chatbot would deliver the suggestions on behalf of the coach. Each day during the *Training* phase, the chatbot reminded the user that the coach had left some journaling suggestions at the beginning of the suggestion session. This system design for the HC condition simulates a scenario where a real coach gives guidance to users through a chatbot. We are interested in how users would react to the guidance mediated by a chatbot. To design a coach's persona, we selected a human name (Sue) and used a coach icon to remind the participants that the guidance was given by the coach (Fig. 1).

Note that all participants received exactly the same suggestions/guidance from the chatbot system irrespective of the condition. The differences between the two conditions are the switching agent icons in the interface (Fig. 1) and the introduction of the coach in the *Suggestion* session.

#### 4.1.1 Three-Phase Study Design.

To observe and understand how the designs affected users' behavior and experience over a period of time, we designed a four-week study. The first day of the four-week study was treated as a practice day to familiarize the participants with the chatbot system, and the remaining 27 days were divided into three equal segments/phases of nine days each (Fig. 2): *Warm-up*, *Training*, and *Free-will*.



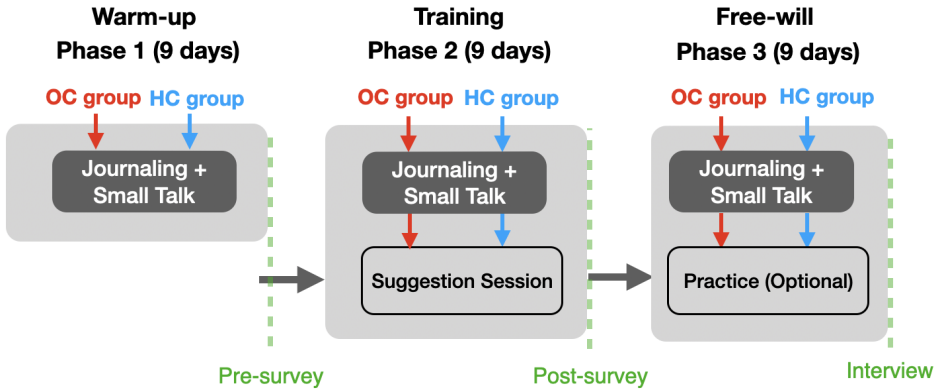


Fig. 2. Study design - The study consisted of three phases: *Warm-Up*, *Training*, and *Free-will*. **OC** represents the group with Only Chatbot, and **HC** represents the group with Human support and Chatbot. Human supporter (coach) appeared only during the *Suggestion* session (in *Training* phase) to deliver suggestions. Other chatting sessions were kept the same for both groups.

In **Phase 1: Warm-up**, each chatting task commenced with a *Journaling* session, followed by a *Small-talk* session. This phase was utilized as a warm-up to familiarize participants with chatting with the chatbot and to remove novelty effects.

In **Phase 2: Training**, a *Suggestion* session was added after the journaling and small-talk sessions (Fig. 2). In the *Suggestion* session, either the chatbot (OC condition) or the coach (HC condition) gave the participants suggestions and guidance to improve their journaling skills and learn new journaling skills. This phase was to investigate users' compliance with the journaling suggestions and measure their perceived interaction (engagement, trust, and intimacy) as well as its effect. On the first day of Phase 2, the chatbot told the **OC group**: "From today, I am going to give you some guidance to learn new journaling skills, which could help you (1) gain a better understanding of your own mental-health status; and (2) help you to improve your happiness and well-being." In the **HC group**, the same comments were attributed to a coach called Sue, who was introduced by the chatbot as follows: "I am going to introduce my colleague, Sue, to you. She is a coach to guide you to learn new journaling skills. She will leave some suggestions for you every day because she cannot always be online to chat." To remind the HC participants that journaling suggestions were provided by the coach, at the beginning of every *Suggestion* session, the chatbot noted that they had suggestions from Sue. Furthermore, whenever the chatbot delivered Sue's message (i.e., the suggestions), it switched the icon to the coach's agent icon and showed her name on the interface. Once the chatbot finished delivering Sue's suggestions, the icon switched back to its original chatbot icon. During the *Training* phase, guidance messages - from the coach to the HC participants and from the chatbot itself to the OC participants - constituted 35% - 40% of the conversations in the (*Suggestion* session).

Lastly, in **Phase 3: Free-will**, no further journaling suggestions were given. Instead, the chatbot encouraged the participants to practice the journaling skills (Fig. 2, *Practice (Optional)*) that had been shared with them during the *Suggestion* session, though they could skip such encouragement without receiving any penalty. The purpose of this phase was to enable the researchers to gauge differences in how our OC and HC conditions affected the participants' journaling practices when following guidance was not required.



## 4.2 Conversation Sessions Across the Three Phases

The conversation flow design was inspired by existing chatbot designs for healthcare (e.g., Woebot<sup>1</sup>), in which the conversation usually begins with a greeting and then proceeds to in-depth conversation. We extended such designs by adding small-talk, which has been shown to enhance users' engagement with and self-disclosure to chatbots [35, 50, 61].

**Journaling Session** - Because of journaling's various documented benefits to personal mental well-being, we designed this type of session - in which users are asked about their moods and the reasons for them - to occur every day. This regularity was also expected to reinforce the participants' awareness that the chatbot was focused on journaling, and not random chit-chatting. Accordingly, our chatbot prompted the participants to focus their journaling on five topics: their mood, experiences, gratitude, stress, and anxiety. Specifically, after an opening greeting, it asked the participant to summarize his/her mood and its causes (e.g., *"Could you let me know what happened to make you feel this way?"*). After any necessary follow-up questions, the chatbot would continue by asking three to five journaling-relevant questions. In such sessions, the chatbot primarily "listened," i.e., gave simple, general responses like *"I understand"* and *"Okay"* or encouraged participants to elaborate on their own answers.

**Small-Talk Session** - Previous research [7, 35, 61] has reported that small-talk (social chat) with a chatbot can improve users' experience of the chatbot system and their performance within it. Our chatbot was designed to engage in self-disclosure, and shared its personal stories in the small-talk sessions. This design was motivated by previous studies' findings that a chatbot's self-disclosure had positive effects on its users' self-disclosure depth [35], and that mutual self-disclosure could enhance users' positive expectations and motivation, and strengthen the therapeutic bond [19]. The conversational topics used in our small-talk sessions were adopted from previous studies [2, 35] and included feelings, thoughts, and information. The same small-talk topics and responses from the chatbot were received by both the OC and HC groups.

**Suggestion Session** - As noted earlier, the *Suggestion* session was only deployed in the *Training* phase, and both the OC and HC groups received exactly the same suggestions and guidance but from the chatbot (OC) and the coach (HC) respectively. The suggestions were adopted from pre-existing journaling materials aimed at improving people's journaling skills, with the wider aim of boosting their well-being [13, 59]. Based on the findings of research in positive psychology [54], some interventions can easily be implemented through typing or writing. For example, Gratitude Journaling [13, 39] is an effective skill/tool for the in-depth expression of appreciation to others, as a path to enhancing one's own well-being and self-reflection.

Our chatbot system was designed to facilitate that prior studies' guideline [39, 47, 51, 55] to build users' journaling skills. For instance, to aid acquisition of the "gratitude journal" skill, the chatbot first explained the benefits of having that skill, and asked its users to think of someone to whom they were grateful. Then, it gave the instructions: *"Let's imagine you are writing a gratitude letter. Take a few minutes to express your gratitude for what this person has contributed to your life. Please directly address the person throughout the letter."* Next, the chatbot encouraged the participants to give more specific and detailed information: *"Describe specific things that this person has done that made you grateful,"* and *"Describe how this person's behavior has affected your life."* Finally, the chatbot asked the participants to wrap up: *"End the letter in a way that identifies it is from you."*

As such, the chatbot's role was to keep guiding users toward the next step, while giving instructions intended to stimulate deep disclosure of their thoughts and feelings. As briefly noted above, the difference between the OC and HC groups was that the latter's instructions - though

<sup>1</sup><https://woeobot.io/>



identical to those provided directly by the chatbot to the OC group - were acknowledged to have been written by a human coach, and the chatbot was portrayed as her intermediary.

The participants would see the coach's message separately from the chatbot's messages in the *Suggestion session*, under their own chatting partner image (Fig. 1, Right). Only one skill (i.e., gratitude journaling [39, 55], expressive writing [47, 51], or the best possible self exercise [45, 56]) was delivered to any participant on a given day, and the participants learned that skill over three consecutive days.

### 4.3 Interface and Implementation

Fig. 1 shows our chatbot interface. Because of its similarity to commercially available messenger applications, the participants readily learned how to use it. They were allowed to give free-text responses to almost all of the system's questions, though sometimes, it provided a few options for them to choose from. For example, when the chatbot asked about a user's mood, it would also show him/her a list of words that could be used when answering. For the chatbot's appearance, we adopted a bot figure (Fig. 1, Left). However, for the group with human support, when the chatbot delivered the human coach's suggestion, a female figure named 'Sue' appeared to visually mark that the suggestions were from the coach (Fig. 1, Right). To prevent users from having a biased view on the coach, we avoided giving them a profile or a portrait of the coach; we instead gave an abstract figure, which is sufficient to remind the users that the messages were from the coach.

Our chatbot was developed using Manychat<sup>2</sup> and Google Dialogflow<sup>3</sup>. The former enabled us to generate daily chatting tasks, to monitor whether the various participants had completed those tasks, and, where necessary, to issue task reminders. The daily chatting tasks included predefined questions and responses from the chatbots, which allowed us greater control of the experimental conditions than would have been possible otherwise. The purposes of integrating Dialogflow, on the other hand, were to increase the naturalness of user-chatbot conversations and to handle users' exceptional questions and requests. If the chatbot detected that a participant became stuck three times within the same chat, it moved to the next conversational topic.

Incorporating natural language processing (NLP), Dialogflow enabled the chatbot to give plausible responses to a wide range of questions from the participants, such as, "*Can you help me?*" and "*How are you?*" Additionally, if a participant said "*I feel stressed today,*" the chatbot's response might include a follow-up question such as, "*Could you let me know why you feel stressed?*" along with its main reply. And, when participants asked questions that the chatbot did not expect and/or could not answer, e.g., regarding human activities such as schooling or diet, Dialogflow helped it process them by providing either simple, naturalistic answers, or requests to rephrase the question or refocus on the task at hand. Besides, this NLP engine allowed the chatbot to handle an additional social chat when participants prompted the chatbot after finishing their daily chatting task. To facilitate natural interaction and transition between the chatting tasks, we designed "intents" in Dialogflow, such as, replying to greetings, replying to users' positive or negative comments, and courtesy. We also designed intents that encourage a user to clarify questions and help them redirect back to the main conversational flow.

### 4.4 Participants

We used social media and a university's electronic bulletin board to recruit participants who met the following criteria: they 1) were above age 18; 2) were able to read, write and speak fluent English; and 3) had a score lower than 13 on the Kessler Psychological Distress Scale [49], meaning that they

<sup>2</sup><https://manychat.com>

<sup>3</sup><https://dialogflow.com>



did not have an urgent mental-health issue. We excluded the latter group because the chatbot was designed to ask its users to share personal experiences that might be negative, and this was deemed to be a potential risk to the well-being of people who were already mentally distressed. In the recruitment stage, we disclosed the duration of the study (four weeks), along with the participants' right to drop out at any point, and each participant's option to attend a follow-up interview at the end of the study.

The 35 participants we recruited ranged in age from 20 to 29 ( $M = 24.74$ ;  $SD = 1.96$ ), and none of them reported having any mental illness. All were graduate or undergraduate students. We divided them into two groups: the OC group (Participant (P)1-17), which received suggestions from the chatbot, and the HC group (P18-35). Both groups were balanced by K-6 score (average OC  $M = 7.97$ , HC  $M = 8.3$ ) and gender, in light of prior research findings on the potential impact of both mental status and gender on self-disclosure [12, 23, 49]. The final composition of OC was 10 females and seven males, and of HC, 11 females and seven males. All participants were familiar with messenger platforms. After the four-week period of interacting with the chatbot, 34 participants attended face-to-face individual interviews, each of which lasted between 45 and 60 minutes.

The participants were paid \$230 USD each, of which \$20 was for the final interview, and the remainder for their participation (chatting tasks and two surveys) in the study *per se*. The same amount was paid regardless of the conditions of participation. On average, we paid \$7.50 for each chatting task (28 chatting tasks in 4 weeks) which reflects the local part-time rate. The participants were told that their financial compensation was contingent upon their completion of at least four chatting tasks during each of the four weeks of the study. If a participant could not meet this requirement, they would get compensation proportionally. However, the participants were also told that they would be able to get the full amount as long as they met this minimum requirement - they would not have to answer all the questions or follow the guidance if they didn't want to, and would not receive any penalty.

#### 4.5 Procedure

All participants were asked to attend an initial meeting in which the researchers helped them set up the chatbot, either on a mobile phone or another device of their choice. In the same meeting, the participants were informed that their conversations with the chatbot would be recorded and analyzed by the research team; that they had the right to refuse to answer questions and to decline to comply with instructions from the chatbot; and that they could withdraw from the experiment whenever they liked. They then signed consent forms. This was followed by a 10-minute chatbot practice/familiarization session. Experimental-group assignments (i.e., to OC or HC) were not discussed with the participants at any time, and they were instructed not to discuss their respective chatbot interactions with one another until after the experiment was finished.

The chatbot prompted the participants when a new chatting task was available. Full access to the chatbot was allowed during 18 hours of each 24-hour period, i.e., from 6 p.m. until 11:59 a.m. the following day. This time-window was selected because, although the researchers preferred interaction to occur in the evening, to ensure that the participants' reactions to the day's events were fresh, they acknowledged that some participants might only have the opportunity to engage with the chatbot in the morning. When a participant accessed the chatbot between noon and 5:59 p.m., it would provide only a simple social chat designed not to affect participants' perceptions of the chatbot or the experiment.

To examine the impact of whether suggestions were given directly by the chatbot or by the coach, we conducted two surveys, one at the end of Phase 1, and the other at the end of Phase 2 (Fig. 2). The details of the survey and other data-collection instruments and approaches will be provided



in the next section, below. At the end of the study, all participants were invited to a face-to-face interview. This research was reviewed and approved by our institutional review board.

#### 4.5.1 Conversation Logs.

We collected all participants' chatbot conversation logs, and compared them across the OC and HC groups. To assess how closely the participants followed the system's suggestions, we focused on changes in the depth of their self-disclosure, which the acquisition of journaling skills has been shown to deepen. Previous research has also used depth of self-disclosure as a metric of users' quality of responses to a chatbot [66] and of their trust in a chatbot system [34]. We had two raters code the log data independently according to the three categories proposed by Barak and Gluck-Ofri [4], i.e., information, thoughts, and feelings, each of which is further subdivided into three levels according to the sensitivity of the information disclosed, as shown in Fig. 4.

Specifically, the raters deemed a user's statement in the log data to be information if it provided factual details about him/her; as thoughts if it expressed his/her personal opinion about events, personal appearance, or intimacy; and as feelings if it expressed an emotional reaction to events, people, or behaviors. The coding of each statement included rater-assigned scores on a scale of 1-3 in each of these three categories, to reflect that they are not absolute or mutually exclusive. Fig. 4 shows the examples. The raters practiced rating a large subsample of users' statements and discussed differences until they reached a consensus, before rating the remainder of the data. A final inter-rater reliability of 89% was achieved.

To analyze how the two chatbot versions, OC and HC, had influenced the participants' depth of self-disclosure, word count, and responsiveness to journaling suggestions, we performed a mixed-model analysis of variance (ANOVA), followed by a Tukey's test of significant difference. For this purpose, we treated the chatbot's suggestions as random effect; groups (OC and HC) and experimental day (Phase 2: 9 days) as independent variables; and self-disclosure levels as the dependent variables.

#### 4.5.2 Survey.

In the surveys administered before and after the *Training* phase (Fig. 2), three constructs were measured to focus on participants' perceived interaction and to facilitate comparison of the effects of the OC and HC conditions. We also evaluated participants' perceived benefits of the interaction by measuring their self-reflection and self-awareness levels. Each construct is described in turn below.

**Perceived Trust:** We measured perceived trust because it is crucial to individuals' decisions to engage in self-disclosure and accept the suggestions, regardless of whether their interlocutors are humans or machines [34, 62]. Our nine measurement items for this construct were adapted from prior literature on perceived trust in computer agents [5, 33, 34], and answered on a seven-point Likert scale ranging from 1=strongly disagree to 7=strongly agree. Example items from this dimension included, "*The chatbot provides me with unbiased and accurate suggestions*" and "*I can trust the information provided by the chatbot.*"

**Perceived Engagement:** The participants were asked to self-report their perceived engagement with the chatbot, as such engagement could have influenced their responsiveness to journaling guidance, and thus their performance when using the system. Additionally, prior studies have suggested that human support probably increases users' engagement, and that loss of engagement is a common obstacle to self-guided treatment [41]. The 12 measurement items for our surveys' engagement dimension were adapted from prior literature [44]. Examples from this dimension included, "*I felt interested in this experience*" and "*The time I spent using the chatbot just slipped away.*"

**Perceived Intimacy:** People may develop intimacy with a chatbot through mutual self-disclosure behavior [6, 35], and they may also perceive intimacy and companionship with a computer agent



Bot:	Let's practice gratitude journaling talk. Choose someone in your life who has been helpful and kind to you, but to whom you haven't had an opportunity to express your gratitude.
User:	Okay
Bot:	Who is the person that makes you feel grateful?
User:	My Grandma.
Bot:	So, directly addressing that person, describe specific things that they've done that made you grateful.
User:	Since I was a child, I never really expressed my gratitude to her. She babysat me when I was an infant and when Mom had to go to work. I was so lucky to have been babysat by a close family member and to get their whole attention, instead of going to a day-care center and being looked after by a stranger.
Bot:	Okay, how has this person's behavior affected your life?
User:	I heard her life story, so I knew that she went through a lot of hardship when she was young. She gave me a lot of life guidance and taught me that we should be generous about helping others when we are able to.

Bot:	Not everyone has had a single big trauma, but most people have had major conflicts or stressors, and you can write about these as well. What is it you feel guilty or stressed about?
User:	Since childhood, I have faced all things by myself, and I will bully and blame others because of envy and even jealousy. I also want love and care and wonder why others can have them, but I can't. This jealousy makes me feel strange.
Bot:	How do you feel right now about this? What feelings and emotions are you experiencing?
User:	When I was little, I would gain satisfaction by bullying classmates, making them cry and making them feel pain. Sometimes these kinds of impulses still appear, but not to bully others.
Bot:	What thoughts show up regarding this situation?
User:	I think it's bad, it's unhealthy, but I don't know how to build good relationships with people.
Bot:	How would you summarize the lesson you learned from this experience?
User:	I often use my past experience to limit myself. Those experiences of hurting others always restrict me, so I don't know how to establish the right relationship with others, or even what such a relationship would be like. But now, I know my past experience is not correct. I can learn from it now.

Fig. 3. Samples of the participants' dialogues. Note: We removed some information from the users' responses for privacy reasons.

when interacting with it over periods of time [27, 35, 40]. Meanwhile, the change in the chatbot's role from a listener in Phase 1 to a coach in Phase 2 may have a negative impact on users' perceived intimacy with the chatbot. Hence, we measured this construct to see if intimacy between each user and the chatbot evolved over the course of Phase 2, and whether/how this construct affects users' depth of self-disclosure to the chatbot system. The six items in this dimension were adapted from a prior study [33]. An example item from this dimension included, *"I felt close to the chatbot."*

**Self-reflection and Self-awareness:** Finally, we measured user's self-reflection and self-awareness because these constructs are potential indicators of the effectiveness of the journaling guidance and our system. Indeed, the journaling suggestions provided by the chatbot/coach were aimed at building participants' self-reflection [21] and self-awareness [20]. We measured users' self-reflection and self-awareness by utilizing a self-reported questionnaire. The 12 measurement items for self-reflection and three items for self-awareness were adapted from prior research [20, 21]. Examples from the two dimensions included *"I frequently take time to reflect on my thoughts," "Right now, I am conscious of my inner feelings,"* and *"Right now, I am reflective about my life."*

We conducted mixed-model ANOVA to analyze the survey results, with the dependent variables being self-reported engagement, trust, intimacy, self-awareness and self-reflection. The two independent variables were group membership - i.e., of OC vs. HC - and time (Fig. 2): i.e., before Phase 2 (Pre-survey) vs. after Phase 2 (Post-survey). Mauchly's test was used to verify that the assumption of sphericity was not violated (Sig. > .05), and Greenhouse-Geisser correction was used to adjust for lack of sphericity.

#### 4.5.3 Interview.

Our semi-structured interview protocol focused on the participants' chatbot experiences, including their daily practices of using the system, their engagement, and their impressions of the chatbot. Follow-up questions covered if/how their attitudes and impressions had changed since the start of the experiment. To capture differences in how the interviewees responded to the chatbot's (or coach's) specific guidance, we asked them to describe their feelings about those suggestions,



including if they felt they were worth following; if they felt comfortable about receiving them; what they learned after following/ignoring them; and if such feelings changed over time.

In addition, we asked the participants how much effort they expended in learning from the system during Phase 2; and how much they disclosed to the chatbot/coach when practicing the journaling skills they had learned. As well as their general impressions of the chatbot/coach, we asked them how they defined their relationship with it/her; if they trusted it/her, and why; and if they perceived themselves as having become dependent upon it/her when learning and practicing journaling skills. Finally, we asked interviewees to reflect upon whether their participation in the study as a whole had affected their self-reflection or self-awareness in daily life, and whether they were still using any of the journaling skills they had been taught during the experiment. Each interview was audio-recorded and transcribed for further analysis.

Thematic content analysis, which involves iteratively reviewing and labeling interviewees' responses with emerging codes, was applied to all the interview data by two raters working independently. The raters' coding results were then compared, and possible revisions were discussed. This cycle was repeated until the coding scheme was deemed satisfactory by both raters, and inter-rater reliability had reached a reasonable level ( $> 91\%$ ).

## 5 RESULTS

To answer the three research questions, we present the results following their order. First, RQ1 is answered by analyzing conversational logs in the *Suggestion* session. Second, RQ2 is answered by the pre- and post-survey of users' perception, and the interview results are included to explore the reasons for causing the experience. Finally, RQ3 is answered by counting the number of participants that voluntarily practiced the journaling suggestions to understand the lasting effect. The interview is also involved in extending understanding. Fig. 3 shows two participants' sample dialogues with the chatbot system.

### 5.1 Effects of Chatbot Designs on Users' Journaling Behaviors (RQ1)

To explore RQ1, regarding how users responded differently to versions of the same chatbot with (HC) and without (OC) a human expert (coach) in the same communication channel, we analyzed conversation logs from the suggestions sessions during the *Training* phase (Phase 2). Since the suggestions given in either chatbot setting required the participants to disclose and reflect more about themselves than had been the case in Phase 1, we measured the quantity and depth of participants' self-disclosure. The statistical results are summarized in Table 1.

On average, HC participants generated 7.9 messages, and OC participants generated 7.5 messages while practicing each *Suggestion* session's journaling skill. We calculated the word count of participants' responses and compared the differences using mixed-models ANOVAs. Results showed that the HC ( $M = 159.41$ ,  $SD = 16.61$ ) group's mean overall word count was significantly higher than that of the OC ( $M = 118.31$ ,  $SD = 17.62$ ) group ( $F(1, 33) = 6.60$ ,  $p < .01$ ). There was no significant main effect of experiment day, and no interaction effects.

#### 5.1.1 Depth of Self-disclosure.

We then coded participants' responses to examine the depth of their self-disclosure. More specifically, first, with regard to *Information*, there was no significant effect of any factor; i.e., neither OC/HC membership nor experiment day significantly impacted how the participants disclosed information to either version of the chatbot. The group averages for informational self-disclosure across all suggestion questions were  $M = 1.7$ ,  $SD = .67$  for OC, and  $M = 1.65$ ,  $SD = .57$  for HC.



Table 1. Statistical results of RQ1: x means no effect. We did not find interaction effects for all measurements

		Group (HC v.s. OC)	
		Word count	Self-disclosure (Information)
RQ1	Self-disclosure (Information)	HC > OC (F = 6.60, p < .01)	x
	Self-disclosure (Thoughts)	HC > OC (F = 29.6, p < .001)	
	Self-disclosure (Feelings)	HC > OC (F = 12.12, p < .05)	

	Information	Thoughts	Feelings
Level 1	<i>Ice cream cake, it is colorful and taste sweet. (P5, F)</i>	<i>I think that many people want to have a cooking skill. (P14, M)</i>	<i>When I did an internship in a company, I was responsible for a presentation but I did not prepare well. (P10, F)</i>
Level 2	<i>My comfort food is instant noodle because it is easy for me to cook after a busy day. (P6, M)</i>	<i>I think I am a good team player because I have good resilience and flexibility. (P33, F)</i>	<i>I feel frustrated because I am not punctuality. I am often late for a date with my friends. (P24, M)</i>
Level 3	<i>My comfort food is soy milk homemade by my grandma, it's a little sweet and smells good, I felt happy to have it at breakfast when I was a kid. (P18, F)</i>	<i>I'm good at writing and I like it! So, I wish I can be a professional writer in the future. So, I keep improving my language skills. (P8, F)</i>	<i>I feel guilty that I hesitated to help an old man who was hurt, and I did not give any help. I regretted this and kept wondering if this person would have died because of my indifference. (P35, F)</i>

Fig. 4. Samples of participants' responses to the suggestions in Phase 2. The responses were coded to different categories and levels of self-disclosure. Note: Level 1 of Feelings is defined as "No expressing of feelings." [4]

Second, in the case of *Thoughts*, our analysis revealed a main effect of group ( $F(1, 33) = 29.6, p < .001$ ), with the participants in the HC group disclosing more thoughts. However, there was no effect of experiment day, and no interaction effects. The group averages for self-disclosure of thoughts across all suggestion sessions were  $M = 1.89, SD = .13$  for OC and  $M = 2.32, SD = .12$  for HC.

Lastly, there was a significant main effect of group membership on the self-disclosure of *Feelings* ( $F(1, 33) = 12.12, p < .05$ ) - with the members of HC disclosing feelings significantly more than OC members did - but no main effect of experiment day, and no interaction effects. The group averages for self-disclosure of feelings across all *Suggestion sessions* were  $M = 1.91, SD = .11$  for OC and  $M = 2.22, SD = .13$  for HC. Fig. 4 shows sample responses with three levels of self-disclosure from participants. These results show that HC group was found to have given longer responses containing deeper *feelings* and more *thoughts* than the OC group during the *Suggestion session* of Phase 2.

## 5.2 Perceived Interaction with the Chatbots (RQ2-a)

To better understand why the participants in the HC group responded to the system's suggestions more diligently than the OC group did, we investigated inter-group differences in how the participants perceived their chatbot interactions. More specifically, we examined whether and how each group members' perceived engagement, intimacy, and trust with the chatbot changed during the *Training* phase. Such levels were captured through their responses to surveys conducted before and after the *Training* phase. The statistical results are summarized in Table 2. Subsequently, in interviews, the participants were asked detailed questions about their motivations for following (or not following) the system's suggestions; their perceptions of such suggestions; and their impressions of whichever chatbot version they had been exposed to.

### 5.2.1 Perceived Engagement.

The engagement level revealed significant main effects of both group ( $F(1, 33) = 8.63, p < .001$ ) and time ( $F(1, 33) = 4.76, p < .05$ ), but there was no significant interaction effect; with the HC group



Table 2. Statistical results of RQ2: x means no effect. We did not find interaction effects for all measurements

		Group (HC v.s. OC)	Time (Pre- v.s. Post-survey)
RQ2-a	Perceived Engagement	HC > OC (F = 8.63, p < .01)	Pre < Post (F = 4.76, p < .05)
	Perceived Trust	HC > OC (F = 8.28, p < .01)	Pre < Post (F = 16.65, p < .001)
	Perceived Intimacy	x	Pre < Post (F = 13.9, p < .001)
RQ2-b	Self-reflection	x	Pre < Post (F = 11.31, p < .01)
	Self-awareness	x	Pre < Post (F = 7.11, p < .05)

reporting significantly higher engagement than OC, both groups' engagement levels generally increased after the *Training* phase ended. To better understand participants' perceived engagement and motivation to follow the suggestions, we asked participants questions in their interviews to explore their experience, for example, *"Did you engage in practicing the journaling skills with the chatbot/coach? Why or why not? How did this change over time?"*

In the interview, ten **OC group** members said that their engagement with the chatbot increased over time because it sent them useful suggestions and prompted them to accomplish something new every day. For example, one participant (P5, F) said: *"Engagement increased over time because I got used to the chatbot and some of its suggestions were useful. I felt more engaged and it gave me motivation."* Besides, most participants in the OC group felt comfortable about deciding for themselves, on a case-by-case basis, whether to follow the chatbot's suggestions or not follow them. When they did follow them, they generally felt happy and surprised that they had been able to learn something useful from a chatbot. As one put it, *"Although I did not practice those suggestions a hundred percent, I think I am still on track. When you learn something from doing this, you will feel more motivated. So, my next step is to keep practicing them."* (P10, F)

Although practicing new journaling skills seemed to enhance perceived engagement for many participants, four members of the OC group reported that doing so caused them fatigue and annoyance. For example, *"I felt annoyed because some suggestions were time-consuming to carry out. I did not expect that I would have to expend so much effort."* (P17, M) Moreover, those participants expressed discomfort at certain suggestions they received from the chatbot. As one of these interviewees said, *"Sometimes, I felt the chatbot was too bossy, especially when it started to give me suggestions. It was okay when the suggestion seemed useful. But when the suggestion was not useful, or the chatbot prompted me to disclose more, I got a bit annoyed. 'It's just a robot, why is it giving me instructions?' That kind of feeling."* (P3, M) In addition, three participants in the OC group reported that their engagement gradually decreased, due to loss of interest in the chatbot system, i.e., the novelty effect [66]. One of them said, *"I was more engaged in the beginning. This [chatting with a chatbot every day] was a new thing for me. But as I got used to it, my engagement level dropped."* (P14, M)

In the **HC group**, according to their survey responses, the HC group felt more highly engaged with the chatbot system than their OC group counterparts did. Most of the HC group participants reported positive attitudes toward practicing the focal journaling skills, and provided two reasons for this. First, much like the OC group, they felt that the content of the suggestions themselves improved their perceived engagement, for example, *"Although practicing journaling skills was time-consuming, following the suggestions helped me understand myself better. So, though lengthy, the process led me to good results."* (P33, F)

Secondly, more than half (n = 11) of the participants in the HC group highlighted the importance of human support, noting that the involvement of a coach increased their willingness to take suggestions seriously. This was because they thought the coach personalized the suggestions for



them and would monitor their practicing activities on the chat channel. As one explained: *"The suggestions were from a coach. I thought the coach might see my responses and give me further suggestions. So, I was more careful about my responses for the suggestions."* (P24, M)

Nevertheless, the ostensible involvement of a human coach seemed to negatively affect the engagement of a minority of the HC participants. Three of them noted that it increased their expectations: i.e., that they would receive highly personalized suggestions and feedback. Because our study design did not actually offer such features, these users' engagement with the chatbot was deterred. As one of them put it, *"The coach gave me suggestions, but they were general suggestions. In fact, I wanted to have more personalized guidance. I felt the coach did not pay attention on my responses."* (P19, F) Moreover, because the coach only offered general suggestions, several mentioned feeling disappointed that 'she' could not really give them personalized feedback or suggestions. For instance, *"I was kind of disappointed by the low level of her involvement to customize suggestions."* (P26, F)

In addition, the impression of human support caused stress to some of the HC users, who felt the coach was judging their answers. As one of them explained, *"I felt the coach would judge my answers, so I tried to answer the questions as thoroughly as possible, and this made me feel stressed."* (P32, F) While this drove them to implement the suggestions, it also triggered negative feelings when they could not follow the suggestions, for example, *"I felt sorry when I could not follow the suggestions, because I guess the coach put a lot of effort into designing this chatbot to help me."* (P30, F)

### 5.2.2 Perceived Trust.

We analyzed the trust levels reported in the survey, and found a significant main effect of group membership on trust in the chatbot, with the HC group's trust being significantly higher than the OC group's ( $F(1, 33) = 8.28, p < .001$ ). There was also a significant, positive within-group main effect of time on trust ( $F(1, 33) = 16.65, p < .001$ ), and there was no significant interaction effect. This analysis showed that the HC group showed stronger trust in the chatbot with a coach than the OC group's perceived trust. To explore how the chatbot system and human support influenced users' perceived trust in each of those system personae, we asked some interview questions related to trust, such as *"Please describe your overall trust in the chatbot"* and *"Do you feel your trust in the chatbot changed during the study? If yes, how and why?"*

The resulting data from interviews indicated that participants in **both groups** developed some degree of trust in the chatbot system, citing a few common factors that contributed to this dynamic. First, in line with prior studies [34, 50], small-talk seemed to help the participants develop trust in the chatbot while building relationships with it. For example: *"I could not trust this chatbot at the beginning of the study because it did not understand me [...] But this feeling of discomfort lessened over time. I felt like we knew each other to some extent because we had some small talk for different things. So I gradually felt I could trust the chatbot more."* (P11, M) Second, four of the 35 participants reported that their trust in the system was rooted in their trust in the research team. As one participant explained, *"I had high trust from the beginning of the study, and this did not change at any point. I believed the research team, so I could trust in this chatbot system."* (P33, F)

In the **HC group**, given that all the journaling suggestions they received allegedly emanated from a human coach, the HC participants' impressions of the chatbot *per se* did not change much across Phase 1 and 2. Most of these participants mentioned that the chatbot's role changed slightly, i.e., from a conversational partner to the coach's assistant, but none of them reported feeling annoyance toward the chatbot. As one of them mentioned, *"After the coach came in, the chatbot became like an assistant to help the coach to deliver suggestions. I can trust the human coach which also enhanced my trust in the chatbot system."* (P28, M)



Furthermore, the presence of a human coach notably increased perceived trust in the chatbot and its suggestions. In the interview, 13 of the 18 HC group members indicated that the introduction of the coach enhanced their trust in the chatbot. It also seemed to have a strong impact on the participants' perceptions of the usefulness of the suggestions they received. For example, one interviewee said, *"I felt suggestions given by a coach would be reliable. For some suggestions, I wasn't sure if they would be effective. If they hadn't been from a coach, I would have been more suspicious and not followed them."* (P30, F) One of the HC participants explained that *"If the chatbot itself gave me some suggestions, I would feel like I was getting the suggestions from a website or a book. The suggestions might still be reliable. But, when the suggestions were from the coach, it was different. I felt a human coach was more careful about the suggestions and my answers."* (P25, F) Thus, the presence of a coach appears to have enhanced users' confidence in the suggestions.

### 5.2.3 Perceived Intimacy.

The participants' self-reported intimacy revealed no main effects of group and no interaction effects; however, the main effect of time was significant ( $F(1, 33) = 13.9$  and  $p < .001$ ). The analysis revealed that this result was driven by both groups, i.e., that the members of both HC and OC groups perceived significantly higher levels of intimacy with the chatbot at the end of the *Training* phase.

For **both groups**, most of the participants confirmed that they perceived their levels of intimacy with the chatbot as having increased over time, but many in the HC group stressed that such growth in intimacy was not influenced by the introduction of the coach. This seems to be borne out by the fact that both groups' explanations of why their intimacy increased were broadly similar. For example, *"The chatbot was not like general mobile APPs. The chatbot seems to have life, so I became closer to it. It reminded me and prompted me to finish practice every day; this design made me feel the chatbot cared about me."* (P10, F)

The main such explanation was that they interacted with the chatbot every day; the small-talk sessions in particular were linked to enhanced intimacy, on the grounds that the chatbot disclosed a considerable amount of its own 'personal' information during small talk. This finding echoes those of prior research by Lee et al. [35], that reciprocity of self-disclosure can increase intimacy levels. According to one HC group member: *"My intimacy level increased to some extent because we talked about our own experiences. I felt like we got familiar with each other."* (P18, F)

In the **HC group**, five of the participants indicated that they tried the suggestions out seriously under the coach's guidance due to the feeling of companionship with the chatbot and coach, which brought them benefits leading to promote their further motivation to follow the instructions. As one of the interviewees said, *"I tried following all the suggestions, even those I thought some of them were not useful. And I had positive feelings after following them. [...] I was able to work so hard because I felt I was not alone. There was the chatbot and the coach. I felt I was supported by both of them."* (P27, F)

Nevertheless, two of the HC users specifically noted that they would have preferred to receive some suggestions from the chatbot rather than from the coach, because the former understood them better: *"I respected the coach's expertise and tried to follow her suggestions. But I wished the chatbot itself would give me some suggestions because it had been listening to my stories from the beginning. It knew about my past and my struggles, so it would have been able to work with me more closely."* (P31, M)

**In summary**, the survey results showed that the HC group perceived significantly more engagement with their chatbot, and their trust in the chatbot significantly increased as well; on the other hand, the OC group's engagement with and trust in the chatbot did not show significant changes. In addition, both groups' sense of intimacy with the chatbot significantly increased after



the *Training* phase. Our interview results suggest several factors that contributed to the positive changes of participants' perceived interaction: 1) For perceived engagement, participants who perceived higher usefulness of the suggestions experienced more engagement with the chatbot, whereas those who did not find the suggestions useful gradually lost interest in interacting with the chatbot. Although ten HC participants noted that the involvement of a coach increased their engagement in practicing the suggestions, four HC participants also shared that they felt pressure given the presence of human support. 2) Regarding perceived trust, the presence of a human coach seemed to contribute to the significant increase of the HC group's trust because the participants felt confident with the suggestions from a human coach. 3) Regarding intimacy, the interview results suggested that the participants in both groups felt more intimate with the chatbot as both groups had small-talk with the chatbot over days.

### 5.3 Perceived Benefits of Practicing Journaling through the Chatbots (RQ2-b)

As we have seen, the two groups of participants perceived the system differently; both later reported that it had a meaningful impact on them. There was a main effect of time on self-reflection level ( $F(1, 33) = 11.31, p < .005$ ), but no significant difference was found between the groups (OC:  $M = 4.31, SD = .30$ , HC:  $M = 4.43, SD = .34$ ), and there was no interaction effect. This analysis showed that both OC and HC group members' self-reflection levels increased significantly after learning journaling skills in the *Training* phase.

For self-awareness level, no significant difference was found between the groups (OC:  $M = 4.51, SD = .23$  and HC:  $M = 4.83, SD = .26$ ), but there was also a main effect of within-subject factors ( $F(1, 33) = 7.11, p < .05$ ). Besides, there was no interaction effect. These findings indicate that both groups of participants improved their self-awareness levels after the *Training* phase. Below, we summarize the points about this topic most frequently mentioned by the participants.

**Better Awareness of Own Mental Status** – Although about half of the participants reported that they had never given full attention to their mood or feelings before the study, most seemed to develop an awareness of their inner emotions and their own strengths by participating in the study. As one stated, *"At the beginning of the study, my messages tended to be short, just saying that I was tired or sleepy. But in the end, I noticed that I was talking more and saying 'I'm happy.' I also learned many skills, like gratitude, and explored my strengths. These practices made me aware of many things about myself."* (P22, F)

**Deeper Understanding of Themselves** – The step-by-step guidance provided by the chatbot system also seemed to encourage the participants to reflect on themselves and helped them understand themselves better. As one reported, *"The chatbot helped when it taught me many journaling skills. By trying those skills out, I could understand myself better. By rethinking my past and going through the exercises, I have gained a better understanding of how my current self was formed."* (P15, M) In other words, by following the chatbot's guidance, some users were able to reflect on prior events from different angles and develop new insights about themselves. This echoes a prior study [29] that indicated that writing down their reflections might give users a sense of accountability, where simply thinking about the answers might not.

**Providing Different Perspectives** – Another benefit of acquiring new journaling skills frequently mentioned by the participants was that the chatbot system reminded them to think about the positive aspects of their lives. For example, participant P2 expressed gratitude toward her parents, supervisor and roommate in the *Training* phase reported in the interview, *"This chatbot reminded me that there are many things I can take care of, and that I feel appreciated. I don't have to focus on things that make me depressed and sad all the time."* (P2, F)

**Paths to Better Well-being** – Finally, nine participants reported that chatting with the system brought about some changes in their lives. Although we cannot verify that such changes had



actually occurred, a sample of the relevant comments is provided below. *"The chatbot made me think about my true feelings and reorganized my thoughts. For example, the coach suggested me to send a gratitude letter, so I sent one to my mother. I actually had complicated mixed feelings about her. There were certain things I didn't like about her. I was thankful, and yet I couldn't get rid of my negative feelings about her. But by following the guidance, I came to realize that I didn't understand my mother well."* (P20, F)

In summary, our findings show that the chatbots could effectively deliver guidance for journaling skills, and both groups' perceived self-reflection and self-awareness levels increased after the *Training* phase. The interview data also supports this observation, with most participants agreeing that journaling both encouraged them to reflect on their prior experience and increase their self-awareness.

#### 5.4 Lasting Effect on Participants' Journaling Practice (RQ3)

To gauge the differences in how our OC and HC chatbot versions affected the participants' journaling practices when following guidance was not required, we first calculated the ratio of participants who practiced the skills per day during the *Free-will* phase (Fig. 2). We then compared the mean ratio between the conditions using a *t*-test. Interestingly, the results showed that more OC participants voluntarily practiced the suggestions than HC participants (OC:  $M = .80$ ,  $SD = .08$ ; HC:  $M = .57$ ,  $SD = .12$ ;  $t = 4.15$ ,  $p < .001$ ). To explore the possible reasons for this difference, we asked the participants to explain their motivations for practicing and quitting the journaling skills during Phase 3.

**5.4.1 Reasons of Keeping the Practices.** Across **both groups**, participants reported similar reasons for continuing to practice the skills during the *Free-will* phase. A majority of the sample ( $n = 20$ ) indicated that because they had benefited from practicing the journaling skills, they felt motivated to keep doing so. As one explained, *"When I followed a suggestion from the chatbot, I was excited to know that even a robot could improve my life! It's not the type of advice we can get in our daily lives, so I keep practicing them."* (P15, M)

Some of the participants emphasized that their relationships with the chatbot also encouraged them to keep practicing the taught skills. As one noted, *"I have been asked about my mood [...] every day, so I gradually felt close to the chatbot. Also, I learned a lot from the chatbot, which gave me a good impression of it. I wanted to keep on using those journaling skills."* (P27, F)

Three of the participants indicated that it was helpful when the chatbot reminded them of the journaling skills learned in the *Training* phase. As one of them put it, *"I forgot about some skills quickly. But by reviewing all the journaling skills, my memories were strengthened, and that gave me confidence that I could successfully improve my mental well-being."* (P13, M)

Also, eight participants appreciated the daily prompts and encouragement by the chatbot. One of them said, *"I think it's not a bad thing to be prompted by the chatbot. It would be hard to have a real human reminding me to practice these skills every day. So the chatbot gave me more motivation to continue working on something."* (P33, F) Moreover, many participants were also motivated to follow the system's suggestions by reminders the chatbot sent to them. One of them shared prior experience with cultivating journaling skills; the interviewee said, *"I once tried to do something similar, but I could not keep doing it on my own. However, when the chatbot encouraged me and sent me reminders, I felt motivated to follow its suggestions. This is quite different from practicing alone."* (P18, F)

**5.4.2 Reasons for Quitting Practice.** The participants who did not often practice in the *Free-will* phase gave several reasons for this, which we summarize below. First, many of them felt tired of journaling after the *Training* phase: as one of them stated, *"I wanted to take a short break from what I had been doing."* (P19, M)



Second, practicing journaling skills was deemed optional in Phase 3, and five participants reported seeing this aspect of our system's design as giving them an excuse to skip it. As one mentioned, *"I feel that if I am given too many alternatives I will just give myself excuses. Especially when I was a little busy with other stuff, I might skip the practice."* (P9, F)

Especially, the **HC group** participants gave group-specific reasons for not practicing in the *Free-will* phase. Six of them reported that because there was no longer a coach monitoring them after the *Training* phase, they did not feel it was necessary to keep practicing. One said, *"I stopped practicing the system's suggestions after the coach stopped giving new ones, because no-one was monitoring my responses and I just wanted to skip that."* (P32, F)

Lastly, two members of the HC group noted that practicing the same journaling skills they had learned in Phase 2 a second time would not have brought them new insights, and thus, they tended not to practice in Phase 3. As one of them put it, *"I think those journaling skills are still useful, but less useful than when I practiced them the first time."* (P34, M)

**In summary**, we found that OC group participants practiced the journaling skills more than HC group participants in the *Free-will* phase (Phase 3). Based on the interview, the results imply that this difference may be attributed to the absence of a human coach to monitor and interact with the participants in the *Free-will* phase, and the decrease in perceived benefit from practicing the journaling skills again.

## 6 DISCUSSION

The major contribution of this work is that it has shown the effects of integrating human support into a chatbot system to deliver suggestions. First, our results show that the participants in the HC group tended to value the suggestions during the *Training* phase and tried to follow the journaling suggestions more closely than their OC group counterparts, resulting in longer journaling content with deeper self-disclosure of their feelings and thoughts (**RQ1**).

Second, the HC participants reported significantly higher perceived engagement and trust than the OC participants. Their perception of the available human expert (coach) appeared to boost the participants' perceived usefulness of the journaling suggestions, as well as their actual journaling practices. However, some HC participants also felt like they were being monitored and/or felt pressure to follow the coach's instructions, and some HC participants' expectations of receiving highly customized suggestions from the human coach were not met, leading to negative impressions (**RQ2-a**). Regardless of such difference between OC and HC groups, both groups' participants seemed to perceive the benefits of learning the journaling skills through the chatbots, e.g., their self-reported levels of self-reflection and self-awareness increased after practicing the journaling skills in the *Training* phase (**RQ2-b**). Their interview feedback explained that the step-by-step guidance prompted them to better reflect on their feelings and thoughts.

Surprisingly, we found that the OC participants practiced the journaling skills significantly more than the HC participants during the *Free-will* phase (**RQ3**). This is unexpected because the HC participants exhibited higher engagement with longer and deeper self-disclosure during the *Training* phase. Participants' interview feedback suggested that HC participants lacked accountability for continuing to practice journaling skills because there was no longer a coach "monitoring" in the *Free-will* phase.

### 6.1 The Impact of Changing or Incorporating Chatbot Identities

As noted earlier, the participants did not receive any suggestions during Phase 1 (*Warm-up*) - they received suggestions for practicing journaling skills after they completed Phase 1. This meant that the chatbot, who had been a conversation partner or listener for both groups, switched its role to



either a coaching role for the OC group, or a mediator between the user and the human coach for the HC group.

For the OC group, when the chatbot changed its role from a conversational partner to a coaching role, there were two types of responses. On one hand, some OC participants were happy to learn something new from the chatbot, and most of them did not feel pressure from the chatbot and enjoyed having the company. On the other hand, some participants reported that the chatbot had become bossy and annoying as the chatbot started to give suggestions in the *Suggestion* session. According to the participants, they did not believe the chatbot understood their struggles and real human lives, so they did not feel comfortable with the chatbot's suggestions. They felt that it would be overcorrecting to adopt the chatbot's suggestions. This is different from the participants in the HC group, who received exactly the same suggestions, yet none reported such feelings toward the human coach nor the chatbot.

On the contrary, none of the HC participants reported any complaints about the chatbot's mediating role. Most of them seemed to believe that their main interlocutor was now a human coach: that the suggestions they were receiving were from that person, and that their responses would be monitored by the coach. Comparing the experience of our two groups suggests that a drastic change in chatbot role may cause participants confusion and discomfort. This observation echoes prior research [38] which found that users felt uncomfortable interacting with a conversational agent using the same personality to serve multiple, completely different, roles. Our results suggested that, adding another persona, the human coach in this case, has the potential to mitigate discomfort. That is, the participants valued the system's suggestions differently due to the ostensible involvement of a human coach; while participants in the OC group tended to judge the value of the system's suggestions on their own, most participants in the HC group tended to believe that the suggestions from the human coach were all valuable, even when they felt some suggestions were not useful. In other words, they trusted the authority of the coach, and were willing to follow the coach. This seems to lend support to Mohr et al.'s [41] finding that legitimacy derived from users' perceptions of a coach positively affected their acceptance of that coach's demands.

Meanwhile, our findings also showed side effects of integrating a human coach to deliver guidance. First, although we simulated the coach who gave suggestions asynchronously, some participants in the HC group reported perceiving high pressure, and/or that their interaction with the chatbot was monitored by the human coach in detail. Several participants stated that they felt stressed and apologetic when they did not have time to closely follow the coach's suggestions, due to the time and effort the coach must have spent in crafting them, based on their progress. Additionally, some participants in our HC group expressed high expectations regarding the customization of the system's suggestions and expressed disappointment accordingly. This echoes Kocielnik et al.'s [29] finding that individuals who held high initial expectations about a chatbot tended to be disappointed. However, in our study, this effect was salient only for the HC group.

## 6.2 Compliance & Sustainability of Practicing Skills through a Three-phase Study

Our study findings reveal some challenges for deploying chatbots in real-life situations, such as establishing relationships between the users and the chatbot [8, 35] and sustaining users' interaction with chatbot [66]. Regardless of the positive or negative effects of time, our findings suggest the importance of conducting a longitudinal study when testing chatbot technologies. Below, we discuss our three-phase study setup and the unique value of this design.

Our study started with a nine-day *Warm-up* phase. We set up this phase because prior work suggested that building trust and intimacy toward the chatbot could motivate users to self-disclose more deeply over time [32, 35]. While our interview findings echo previous work regarding the important role of time in establishing social relationships between the participants and the chatbot



[9, 35], our data also showed that some participants experienced a novelty effect [66]: i.e., that they were highly engaged in chatbot conversation at the very beginning, but that this excitement gradually decreased thereafter. Our study results show that the participants' perceptions and attitudes toward the chatbot changed as they interacted with it across the different temporal phases of the experiment.

Through the three-phase study, we were able to find that the HC participants had higher levels of dependency on the chatbot. However, this could have diminished their motivation to keep journaling once it became an optional activity in the *Free-will* phase. According to the investment model of long-term engagement [8], users may stop using an agent if they perceive higher cost and lower benefit. Our findings are in line with the theory [8], insofar as the HC participants perceived higher cost for practicing the journaling skills in the *Training* phase, which demotivated them to continue practicing in the *Free-will* phase. On the other hand, the OC participants had both lower self-disclosure and lower engagement in the *Training* phase than their HC-group counterparts. The lack of pressure to follow system suggestions that the OC participants perceived could result in a lower perceived cost of continuing the practices. Therefore, compared to the HC participants, their willingness of continuing the practices was less negatively impacted in the *Free-will* phase.

### 6.3 Design Implications

Our study uncovered both benefits and drawbacks of integrating human support into human-chatbot interaction. These shed light on future chatbot designs both for practicing journaling skills to improve mental health and for skill training in broader contexts.

The major benefit of integrating expert's advice into human-chatbot interaction is that the expert's involvement resulted in participants' higher engagement. In our context, it was measured by participants' journaling efforts (length and depth of their journaling content) and their perceived interaction from the perspectives of engagement, trust, and intimacy. Prior research has shown that compliance with system requests is an important first step towards users' achievement of further positive behavioral change [43]. Shi et al. [57] found that when users identified a chatbot as a human, they would think that the conversation was more engaging and have better outcomes. Our findings extend our understanding by examining whether users can be persuaded to follow suggestions from a coach even when the suggestions are not customized for individuals and provided asynchronously through a coach's agent; however, the integration of a human coach might lead the participants to think that they were monitored and cause them to feel pressure.

Scholars already suggest that using a chatbot as a coach can guide users toward a healthy lifestyle and activities, e.g., [18, 29, 37]. Additionally, coordinated with other technologies such as physical sensing and machine learning, future virtual agents, e.g., Alexa and Google Assistant [61], could automatically track users' behaviors and introduce proper expert services. When human experts, e.g., professional coaches, counselors and healthcare providers, are scarce or have limited availability, our proposed HC design could be adopted to help a human expert deliver their suggestions to the clients asynchronously, and the chatbot could play a role to help collect and track the clients' information to assist human experts in offering more persuasive guidance than that given only through chatbots. Future designs could also consider defining the human expert agent as a crowd-powered expert to flexibly attribute limited human expert resources among users.

Finally, we found that the chatbot versions in the OC and HC conditions had specific advantages and disadvantages when it came to delivering guidance. Future practitioners and designers could consider the factors, e.g., switching chatbot identities and sustainability of skill practices, when using chatbots for different training purposes. For example, if the specific skill requires close compliance in a short period of time, incorporating human experts may be more effective in the training phase. Conversely, if the training requires users' long-term engagement and a light touch



with the experts, it may be preferable to have a companion chatbot lead the interaction. More varied design options need to be further evaluated in the specific application contexts.

#### 6.4 Ethical Considerations of Incorporating Human Support in a Chatbot System

Chatbot works need to address ethics and privacy issues carefully [24, 30, 34]. For example, because chatbot use can result in users' self-disclosure of sensitive topics [30, 34], whether to disclose it, when, and precisely to whom remain difficult to answer. This is especially challenging to answer if the goal is to improve mental health. If a chatbot introduces an expert (a coach or counselor) in the middle of the study, it may raise privacy and transparency concerns because it may imply that past conversations between the chatbot and users are transferred to other parties. Our participants did not specifically express this concern in our study, which may be attributed to the social presence and expertise of a coach and our research setting. Nevertheless, this concern may be heightened when the chat topic is sensitive, and it is an important issue to explore in future research.

Most of our participants tended to trust the system's suggestions. In the HC group, introducing the human coach reinforced the participants' perceived trust and raised their expectations on the coach for giving personalized suggestions. Although our research did not aim to deceive the users, our study findings revealed that integrating a human expert (coach) might lead to participants' unrealistic expectations of a chatbot's intelligence and efficacy. Hence, when deploying a chatbot system for healthcare purposes, the transparency of the mechanism should be informed. In addition, although the journaling suggestions our chatbot systems offered were evidence-based [39, 47, 55, 56] and had no potential to cause harm, 'over-trust' could lead to harm if a future system were not working properly [67]. Our research is a good start to inform how future studies should be aware of these risks when using conversational agents to deliver guidance.

### 7 LIMITATIONS AND FUTURE WORK

There are several limitations of this work. First, to keep the two groups comparable, we controlled the chatbot responses when providing suggestions for the users. There was not an actual human person monitoring the participants and personalizing suggestions for them in this study. Thus, our study design did not perfectly match the participants' expectation of customized expert support. However, the result could also reveal how users might react when a system does not work as they expect. Still, it will be beneficial to further examine the impact of offering personalized guidance in further research. Future research could also evaluate how synchronous communication with real human experts may impact user experience. In addition, we did not explore the different designs of chatbot personas (i.e., name and icon) to represent the coach in this work. Future work could be focused in this direction by examining the effect of controlling the presentation of a human expert's persona.

Second, we gave a relatively easy task to the participants to learn (journaling skills). We used the depth of self-disclosure and word count to measure how the users followed the suggestions. Because journaling needs users to disclose themselves, the depth of self-disclosure level could reflect how users put effort into responding to those suggestions. Users' daily experience and behavior might be affected over time. In our study, we did not focus on behavioral change.

Finally, although we believe that the differences between the OC and HC groups were not affected by compensation as our OC and HC participants were compensated at the same rate, the engagement level of the participants taken as a whole was surely affected by compensation. Future work should consider deploying the system without compensating the users for a longer-term span in different contexts.



## 8 CONCLUSION

In this paper, we explored chatbot designs with and without integrating human support for delivering guidance to users, along with a three-phase study we conducted to examine the impact of human expert support on human-chatbot interaction. We focused on investigating the impact on users' experience and efforts involved in following the guidance of journaling skills. Our findings show that the chatbot integrated with human support (HC) enhances users' self-disclosure, perceived engagement and trust in the *Training* phase; however, it might also make people feel pressured. Although the group without human support (OC) showed less engagement in the *Training* phase, they remained interested longer in practicing journaling skills. Our research extends the understanding of human-chatbot interaction by providing insights on future chatbot designs with the support of human experts to deliver guidance in broader contexts.

## 9 ACKNOWLEDGMENTS

This work is supported by Grant for Scientific Research (A) 17H00771 from Japan Society for the Promotion of Science (JSPS). We also thank the Faculty Start-up Funds provided by the School of Information Sciences at Illinois for supporting this research. Finally, we thank Prof. Takashi Kudo for early feedback in the study, and we are thankful for all the anonymous reviewers whose feedback helped us improve the paper significantly.

## REFERENCES

- [1] Gerhard Andersson and Pim Cuijpers. 2009. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cognitive behaviour therapy* 38, 4 (2009), 196–205.
- [2] Arthur Aron, Edward Melinat, Elaine N Aron, Robert Darrin Vallone, and Renee J Bator. 1997. The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin* 23, 4 (1997), 363–377.
- [3] Karen A Baikie and Kay Wilhelm. 2005. Emotional and physical health benefits of expressive writing. *Advances in psychiatric treatment* 11, 5 (2005), 338–346.
- [4] Azy Barak and Orit Gluck-Ofri. 2007. Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior* 10, 3 (2007), 407–417.
- [5] Izak Benbasat and Weiquan Wang. 2005. Trust in and adoption of online recommendation agents. *Journal of the association for information systems* 6, 3 (2005), 4.
- [6] Ellen Berscheid, Mark Snyder, and Allen M Omoto. 1989. The relationship closeness inventory: Assessing the closeness of interpersonal relationships. *Journal of personality and Social Psychology* 57, 5 (1989), 792.
- [7] Timothy Bickmore and Justine Cassell. 1999. Small talk and conversational storytelling in embodied conversational interface agents. In *AAAI fall symposium on narrative intelligence*. 87–92.
- [8] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. 2010. Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence* 24, 6 (2010), 648–666.
- [9] Timothy W Bickmore, Lisa Caruso, Kerri Clough-Gorr, and Tim Heeren. 2005. 'It's just like you talk to a friend' relational agents for older adults. *Interacting with Computers* 17, 6 (2005), 711–735.
- [10] Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. 2017. Towards a chatbot for digital counselling. In *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*. BCS Learning & Development Ltd., 24.
- [11] Ana Paula Chaves and Marco Aurelio Gerosa. 2018. Single or Multiple Conversational Agents? An Interactional Coherence Comparison. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [12] Paul C Cozby. 1973. Self-disclosure: a literature review. *Psychological bulletin* 79, 2 (1973), 73.
- [13] Don E Davis, Elise Choe, Joel Meyers, Nathaniel Wade, Kristen Varjas, Allison Gifford, Amy Quinn, Joshua N Hook, Daryl R Van Tongeren, Brandon J Griffin, et al. 2016. Thankful for the little things: A meta-analysis of gratitude interventions. *Journal of counseling psychology* 63, 1 (2016), 20.
- [14] S Divya, V Indumathi, S Ishwarya, M Priyasankari, and S Kalpana Devi. 2018. A self-diagnosis medical chatbot using artificial intelligence. *Journal of Web Development and Web Designing* 3, 1 (2018), 1–7.
- [15] Wen Duan, Naomi Yamashita, Sun Young Hwang, and Susan Fussell. 2018. "Let Me Ask Them to Clarify If You Don't Want To"-A Clarification Agent for Nonnative Speakers. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.



- [16] David Daniel Ebert, Dirk Lehr, Filip Smit, Anna-Carlotta Zarski, Heleen Riper, Elena Heber, Pim Cuijpers, and Matthias Berking. 2014. Efficacy and cost-effectiveness of minimal guided and unguided internet-based mobile supported stress-management in employees with occupational stress: a three-armed randomised controlled trial. *BMC Public Health* 14, 1 (2014), 807.
- [17] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 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 mental health* 4, 2 (2017), e19.
- [18] Silvia Gabrielli, Silvia Rizzi, Sara Carbone, and Valeria Donisi. 2020. A Chatbot-Based Coaching Intervention for Adolescents to Promote Life Skills: Pilot Study. *JMIR Human Factors* 7, 1 (2020), e16762.
- [19] Marvin R Goldfried, Lisa A Burckell, and Catherine Eubanks-Carter. 2003. Therapist self-disclosure in cognitive-behavior therapy. *Journal of clinical psychology* 59, 5 (2003), 555–568.
- [20] John M Govern and Lisa A Marsch. 2001. Development and validation of the situational self-awareness scale. *Consciousness and cognition* 10, 3 (2001), 366–378.
- [21] Anthony M Grant, John Franklin, and Peter Langford. 2002. The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality: an international journal* 30, 8 (2002), 821–835.
- [22] Brenda Hayman, Lesley Wilkes, and Debra Jackson. 2012. Journaling: Identification of challenges and reflection on strategies. *Nurse researcher* 19, 3 (2012).
- [23] Charles T Hill and Donald E Stull. 1987. Gender and self-disclosure. In *Self-Disclosure*. Springer, 81–100.
- [24] Ayanna Howard and Jason Borenstein. 2019. Trust and Bias in Robots: These elements of artificial intelligence present ethical challenges, which scientists are trying to solve. *American Scientist* 107, 2 (2019), 86–90.
- [25] Delaura L Hubbs and Charles F Brand. 2005. The paper mirror: Understanding reflective journaling. *Journal of Experiential Education* 28, 1 (2005), 60–71.
- [26] Maia L Jacobs, James Clawson, and Elizabeth D Mynatt. 2015. Comparing health information sharing preferences of cancer patients, doctors, and navigators. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 808–818.
- [27] Jinkyu Jang and Jinwoo Kim. 2019. Healthier Life with Digital Companions: Effects of Reflection-Level and Statement-Type of Messages on Behavior Change via a Perceived Companion. *International Journal of Human-Computer Interaction* (2019), 1–18.
- [28] Christina Kelley, Bongshin Lee, and Lauren Wilcox. 2017. Self-tracking for mental wellness: understanding expert perspectives and student experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 629–641.
- [29] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection companion: A conversational system for engaging users in reflection on physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 70.
- [30] Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini, Arianna Manzini, Ilina Singh, and NeurOx Young People's Advisory Group. 2019. Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical Informatics Insights* 11 (2019), 1178222619829083.
- [31] Rohit Kumar and Carolyn P Rose. 2010. Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies* 4, 1 (2010), 21–34.
- [32] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselstein. 2019. Caring for Vincent: A Chatbot for Self-Compassion (CHI '19). ACM, New York, NY, USA, Article 702, 13 pages. <https://doi.org/10.1145/3290605.3300932>
- [33] SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies* 103 (2017), 95–105.
- [34] Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. *Proceedings of the ACM on Human-Computer Interaction* (2020).
- [35] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. *ACM CHI Conference on Human Factors in Computing Systems* (2020).
- [36] Gale M Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. 2017. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI* 4 (2017), 51.
- [37] Kai Lukoff, Taoxi Li, Yuan Zhuang, and Brian Y Lim. 2018. TableChat: Mobile Food Journaling to Facilitate Family Support for Healthy Eating. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–28.
- [38] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Re-Embodiment and Co-Embodiment: Exploration of social presence for robots and conversational agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 633–644.



- [39] Sonja Lyubomirsky, Rene Dickerhoof, Julia K Boehm, and Kennon M Sheldon. 2011. Becoming happier takes both a will and a proper way: An experimental longitudinal intervention to boost well-being. *Emotion* 11, 2 (2011), 391.
- [40] Adam S Miner, Nigam Shah, Kim D Bullock, Bruce A Arnow, Jeremy Bailenson, and Jeff Hancock. 2019. Key considerations for incorporating conversational AI in psychotherapy. *Frontiers in psychiatry* 10 (2019).
- [41] David Mohr, Pim Cuijpers, and Kenneth Lehman. 2011. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. *Journal of medical Internet research* 13, 1 (2011), e30.
- [42] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 72–78.
- [43] Harri Oinas-Kukkonen. 2013. A foundation for the study of behavior change support systems. *Personal and ubiquitous computing* 17, 6 (2013), 1223–1235.
- [44] Heather L O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28–39.
- [45] Nansook Park and Christopher Peterson. 2009. Character strengths: Research and practice. *Journal of college and character* 10, 4 (2009).
- [46] SoHyun Park, Jeewon Choi, Sungwoo Lee, Changhoon Oh, Changdai Kim, Soohyun La, Joonhwan Lee, and Bongwon Suh. 2019. Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study. *Journal of medical Internet research* 21, 2 (2019), e12231.
- [47] James W Pennebaker. 2010. Expressive writing in a clinical setting. *The Independent Practitioner* 30 (2010), 23–25.
- [48] Juanan Pereira and Óscar Díaz. 2019. Using health chatbots for behavior change: a mapping study. *Journal of Medical Systems* 43, 5 (2019), 135.
- [49] Judith J Prochaska, Hai-Yen Sung, Wendy Max, Yanling Shi, and Michael Ong. 2012. Validity study of the K6 scale as a measure of moderate mental distress based on mental health treatment need and utilization. *International journal of methods in psychiatric research* 21, 2 (2012), 88–97.
- [50] Abhilasha Ravichander and Alan W Black. 2018. An Empirical Study of Self-Disclosure in Spoken Dialogue Systems. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. 253–263.
- [51] Lindsey M Rodriguez, Chelsie M Young, Clayton Neighbors, Michelle T Campbell, and Qian Lu. 2015. Evaluating guilt and shame in an expressive writing alcohol intervention. *Alcohol* 49, 5 (2015), 491–498.
- [52] Stephen M Schueller, Kathryn Noth Tomasino, Emily G Lattie, and David C Mohr. 2016. Human support for behavioral intervention technologies for mental health: the efficiency model. *management* 21 (2016), 22.
- [53] Stephen M Schueller, Kathryn Noth Tomasino, and David C Mohr. 2017. Integrating human support into behavioral intervention technologies: the efficiency model of support. *Clinical Psychology: Science and Practice* 24, 1 (2017), 27–45.
- [54] Martin EP Seligman and Mihaly Csikszentmihalyi. 2014. Positive psychology: An introduction. In *Flow and the foundations of positive psychology*. Springer, 279–298.
- [55] Martin EP Seligman, Tracy A Steen, Nansook Park, and Christopher Peterson. 2005. Positive psychology progress: empirical validation of interventions. *American psychologist* 60, 5 (2005), 410.
- [56] Kennon M Sheldon and Sonja Lyubomirsky. 2006. How to increase and sustain positive emotion: The effects of expressing gratitude and visualizing best possible selves. *The journal of positive psychology* 1, 2 (2006), 73–82.
- [57] Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376843>
- [58] Ha Trinh, Ameneh Shamekhi, Everlyne Kimani, and Timothy W Bickmore. 2018. Predicting User Engagement in Longitudinal Interventions with Virtual Agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 9–16.
- [59] Allison Utley and Yvonne Garza. 2011. The therapeutic use of journaling with adolescents. *Journal of Creativity in Mental Health* 6, 1 (2011), 29–41.
- [60] Alastair van Heerden, Xolani Ntinga, and Khanya Vilakazi. 2017. The potential of conversational agents to provide a rapid HIV counseling and testing services. In *2017 international conference on the frontiers and advances in data science (FADS)*. IEEE, 80–85.
- [61] Jinping Wang, Hyun Yang, Ruosi Shao, Saeed Abdullah, and S Shyam Sundar. 2020. Alexa as Coach: Leveraging Smart Speakers to Build Social Agents that Reduce Public Speaking Anxiety. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [62] Lawrence R Wheelless and Janis Grotz. 1977. The measurement of trust and its relationship to self-disclosure. *Human Communication Research* 3, 3 (1977), 250–257.
- [63] Alex C. Williams, Harmanpreet Kaur, Gloria Mark, Anne Loomis Thompson, Shamsi T. Iqbal, and Jaime Teevan. 2018. Supporting Workplace Detachment and Reattachment with Conversational Intelligence (*CHI '18*). ACM, New York, NY, USA, Article 88, 13 pages. <https://doi.org/10.1145/3173574.3173662>



- [64] Rainer Winkler, Sebastian Hobert, Antti Salovaara, Matthias Söllner, and Jan Marco Leimeister. 2020. Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [65] Kieran Woodward, Eiman Kanjo, David Brown, T Martin McGinnity, Becky Inkster, Donald J Macintyre, and Athanasios Tsanas. 2019. Beyond Mobile Apps: A Survey of Technologies for Mental Well-being. *arXiv preprint arXiv:1905.00288* (2019).
- [66] Ziang Xiao, Michelle X Zhou, Q Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2019. Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys. *arXiv preprint arXiv:1905.10700* (2019).
- [67] Jin Xu, G Bryant De'Aira, Yu-Ping Chen, and Ayanna Howard. 2018. Robot therapist versus human therapist: Evaluating the effect of corrective feedback on human motor performance. In *2018 International Symposium on Medical Robotics (ISMR)*. IEEE, 1–6.
- [68] Jin Xu, G Bryant De'Aira, and Ayanna Howard. 2018. Would You Trust a Robot Therapist? Validating the Equivalency of Trust in Human-Robot Healthcare Scenarios. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 442–447.
- [69] Jingwen Zhang, Yoo Jung Oh, Patrick Lange, Zhou Yu, and Yoshimi Fukuoka. 2020. Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet. *Journal of Medical Internet Research* 22, 9 (2020), e22845.

Received June 2020; revised October 2020; accepted December 2020