Delft University of Technology
Electrical Engineering, Mathematics and Computer Science
Department of Interactive Intelligence

Motivating PhD candidates with depression symptoms to complete thoughts-strengthening exercises via a conversational agent

by

Mohammed A. Al Owayyed

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday July 2, 2020 at 14:00.

Thesis committee:	Dr. Ir. Willem-Paul Brinkman,	Associate Professor, TU Delft, Supervisor
	Dr. Myrthe Tielman,	Assistant Professor, TU Delft
	Dr. Annibale Panichella,	Assistant Professor, TU Delft

Abstract

PhD students are usually more susceptible than other people to depression-inducing situations, which are not adequately addressed in eHealth systems. Cognitive behavioral therapy (CBT) is a well-established depression treatment, and one of its therapeutic methods is to change negative thought patterns into positive ones. The new positive thoughts need to be more believable than the old ones, so therapists usually ask clients to complete sets of thought-strengthening exercises. However, psychologists have expressed concerns regarding patients' compliance in doing these exercises, as the majority of clients encounter at least one hurdle that obstructs them from completing the exercises. Therefore, we developed a system aimed at motivating doctoral students to perform thought-strengthening exercises. The system allows the candidates to interact with a conversational agent (i.e., a chatbot) to address motivational barriers related to the exercises and complete the exercises within the same program. We used a double-blind experiment with a mixed design setup that included a healthy general public sample group (n = 174) and three conditions (chatbot support, text support, and no support) to test the defined hypotheses and evaluate the system. We hypothesized that the participants' perceived usefulness and self-efficacy regarding the thought-strengthening exercises would be higher depending on the presentation format (i.e., motivation via text or interactive chatbot) and content (i.e., with or without motivation). The results related to perceived usefulness showed an increase in the conversational agent condition, which showed a significant difference when compared with the no-support condition. As for self-efficacy, various increases occurred in all three conditions; however, no significant difference appeared between them. In conclusion, the conversational agent support system showed the potential to increase PhD students' motivation to complete the thought-strengthening exercises.

Preface

This graduation thesis was completed as part of the fulfilment for my Master's in Computer Science from TU Delft.

The experience of working with passionate people (either during my thesis period or before that, while taking courses) has broadened my perspective, not just on a professional level, but also on a personal level. I used to think that having a skilled teacher would be enough for me to understand the information. However, utilizing this acquired knowledge to reach one's desired goals is what really matters. This actually shows the differences between a good teacher and a good mentor: the latter guides you in applying the acquired knowledge and does not just deliver it. For that, I would like to thank my supervisor, Willem-Paul Brinkman, for his valuable support and guidance during the course of this research and his great facilitating of the weekly research group meetings (and sense of humor). Also, I would like to thank Franziska Burger, who was my daily supervisor for the majority of the thesis period, for her great assistance and feedback. I cannot forget the Interactive Intelligence research group and the useful feedback and knowledge gained during our weekly meetings. Thank you all, from the other master's degree students (David, Siyu, Mitchell, Pascal, Zhang, Zilla, Salim and Jeffrey) to the PhD candidates (Fran and Ding) and the postdocs (Merijn). I also would like to thank my friends here in Delft and back in my country for their support and the fun times we have had together, especially in these unprecedented circumstances during the coronavirus pandemic. Further, thanks to all psychologists and PhD candidates who took part in the focus group sessions and provided their useful opinions.

Finally, sincere thanks goes to my family for their continuous support: to my parents, whose wisdom and kindness always served me well, and to my wife, Hanan, to whom I am grateful for encouraging and standing up for me during my studies and life.

Mohammed A. Al Owayyed
Delft, June 2020

Contents

1

Background and Motivation

To the best of our knowledge, no existing computer-based depression interventions are tailored to PhD candidates, even though these individuals are more likely than the public to be exposed to depression-inducing situations [34]. The common obstacles that PhD students face during their studies can generate negative thinking patterns. Replacing these negative patterns with healthier ones is one prevalent method of treatment for such situations within the framework of cognitive behavioral therapy (CBT). This process, however, typically requires great effort; a person can believe in new thoughts but still feel that these thoughts are not as applicable to their lifestyle or experiences as the old (negative) thoughts [59]. Therefore, training and strengthening the newly formed thoughts is an important aspect of CBT, as it helps enforce the new thoughts in one's mind. This research aims at building a prototype of a conversational agent intervention for motivating PhD candidates to perform thought strengthening exercises.

1.1. Underlying Theory

According to the World Health Organization (WHO), depression and anxiety are the most prevalent mental illnesses in the world [58]. In addition, researchers have reported that university students have higher depression levels than the general public [98] [43]. Many of these studies were focused on depression symptoms in undergraduate students. The Graduate Assembly of the University of California, Berkeley, recently published a report on the happiness and well-being of that institution's graduate students [95]; per this report's depression questionnaire, about 47% of PhD students scored at least 10 out of 30 on the depression scale, thus reaching the threshold for depression. Furthermore, in its 2015 annual survey, the Central PhD Council of the University of Amsterdam concluded that members of the PhD community were twice as likely to be clinically depressed as the general public, with 36.5% of surveyed PhD candidates scoring above the threshold for depression [34].

CBT is regarded as effective psychotherapy for depression [28]. The therapy consists of two parts: the cognitive part (i.e., cognitive therapy), which aims at improving patients' mood by replacing negative, debilitating thinking patterns with healthier ones and helping to form more realistic and adaptive cognitions [11], and the behavioral part, which aims to get people to engage in activities in order to experience positive rewards. Several researchers have shown that cognitive therapy (CT) has positive mental-health impacts on both graduate and undergraduate students (e.g., in [105]). According to the UK National Institute for Health and Clinical Excellence (NICE) guideline [54], cognitive behavioral therapy should be the first treatment when dealing with depressed patients. The standard approach to thought changing in cognitive therapy is shown in Figure 1.1 [59].

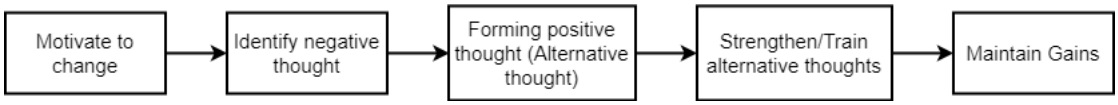


Figure 1.1: "Changing negative thoughts" general pipeline

To motivate the depressed person to change the negative thought, psychoeducation about cognitive therapy is provided. After that, patients think back to an event or situation that upset them and identify the

negative, depleting thought that caused them to feel upset. A positive alternative way of interpreting the situation is sought in collaboration with the therapist. Then, a new thought is formed during the therapy session. This thought needs to be strengthened and trained. Therefore, patients are typically assigned thought strengthening exercises as homework exercises.

According to Kazantzis et al. [68], CBT patients who carry out homework exercises progress better in treatment. Furthermore, performing the exercises is considered as an essential part of the therapy [15] [39], to the point that some researchers claim that it has earned the most empirical attention among all aspects of CBT [97]. However, therapists often struggle with patients' homework compliance. In their experiment regarding CBT homework compliance, Helbig and Fehm [61] found out that around 74.5% of patients encountered at least one compliance problem while performing the exercises.

1.2. Computerized Therapy

Approximately 43 million depression cases worldwide go untreated each year [69]. This treatment gap has various causes, including some patients' inability to participate in treatment, as well as their embarrassment or skepticism about treatment, and financial limitations [69] [115]. Therefore, computerized CBT (cCBT) has been proposed as an alternative solution, as it provides anonymity, constant availability, and ubiquity while also being more affordable than traditional means of therapy [117]. The main concept behind cCBTs is to deliver (aspects of) CBT treatment via computer programs. Overall, research on the effectiveness of cCBT programs have revealed promising results [66]. For example, Day et al. [36] evaluated an internet-based cCBT system targeted toward university students well-being. The session-based program was designed to manage three mood states: depression, anxiety and stress. The authors found an overall improvement in graduate and undergraduate students' mental health from using the system.

Recently, the adoption of conversational computer interfaces (whether text- or voice-based) has increased. Another name for such interfaces is conversational bots, which can be described as artificial intelligences that engage in human-language conversations [100]. Virtual agents such as Apple's Siri and Microsoft's Cortana use natural-language processing to communicate with users and assist them. Moreover, interactive agents are utilized in domain-dependent settings, including in health care, education, commerce, and entertainment [77]. One reason for using such familiar frameworks is to increase the applications' resemblance to real-life textual interactions [82]. In theory, mental health chatbots should be able to replicate some aspects of the therapeutic process while providing a personalized experience that is tailored to the individual patient's issues [14]. Therefore, a few chatbots have been designed to make use of CBT interventions to treat depression. Patients using these bots have shown general improvements in terms of depression and anxiety symptoms [51] [65]. Such conversational bots usually apply CBT principles to improve individuals' mental well-being.

1.3. Research Question

The main goal of this research is to develop a prototype and explore the design of a conversational agent for motivating PhD students to perform thought-strengthening exercises. In other words, alternative positive thoughts are already worked on with a psychologist, but need to be made more believable and automated for the individual by performing homework exercises. Therefore, our research question is as follows:

How can a conversational intervention agent be used to motivate PhD candidates in changing negative thinking patterns?

This research question is divided into three subquestions:

- *What are this conversational agent's **requirements**?*
- *What **design** would meet the requirements?*
- *Can such a conversational agent increase doctoral students' **motivation** toward thought-strengthening exercises?*

1.4. Research Approach

To answer the overall and sub research questions, we followed the situated cognitive engineering approach [90]. The first stage was to examine the current literature to ensure a proper understanding of the problem,

the research domain and the chatbot's potential contributions. In addition, we worked toward deriving the chatbot's full list of requirements by considering the produced scenarios, which were evaluated in conjunction with domain experts in focus groups, and then adjusted accordingly. The main outcome of this phase was a list of requirements that were used as a guideline for the overall functionality of such a system. (Foundation - Chapter 2). These scenarios and requirements were used to propose a design for the prototype by focusing specifically on motivating PhD candidates to perform thought-strengthening exercises and converting these scenarios into dialogues or strategies to use with the chatbot (Design - Chapter 3). After that, an experiment was conducted to evaluate if the developed prototype had the potential to increase PhD students' motivation to completing assignments (Evaluation - Chapter 4). In the end, we conclude our work with a discussion and a summary of the research (Discussion and Conclusion - Chapter 5).

2

Foundation

In this chapter, we worked toward answering the first research question:

*"What are this conversational agent's **requirements**?"*

Determining requirements helped to define what the conversational agent should achieve. We began by investigating the current practices, which mainly included a description of the problem with which a virtual agent can help and an analysis of the stakeholders (operational demands). This was followed by examining past research to gain a clearer perception of how to solve such a problem (i.e., examining human factor knowledge). After that, we explored the related recent technologies to identify strategies or options that could be used in the prototype, and we described how we envision the required chatbot to look (current related technology). This was followed by exploration, for which we conducted two focus groups: one with domain experts and the other with PhD students. The main aim of the focus groups was to gather and refine the final list of requirements, which we ended this chapter with.

2.1. Operational Demands

2.1.1. Common negative thoughts of PhD candidates

There are different factors that can affect PhD candidates' well-being. These factors can be divided into internal factors (the relationship between academic progress and the mental welfare of a student), and external factors, which are concerned with persons, resources and obligations related to the student [109]. In this research, we will focus on the cognitive part (i.e., the internal factors).

Goal-setting theory [79] states that motivation can affect performance and job satisfaction, which aligns with research on PhD candidates' motivation: studies have shown that the higher the PhD students' motivation, the more accomplishments they achieve, thus leading to greater satisfaction (e.g., [19] [40]). One issue that could deteriorate motivation is self-worth. For PhD students, the most notable concern related to self-worth is their sense of entitlement (are they entitled to their current positions?) [41]. This is affected by facing and overcoming academic hurdles, such as critical feedback from reviewers on a paper or praise from a supervisor [81]. The result might be an "elation and depression cycle," which means a fluctuation between feeling capable and frustrated [109]. Another factor moderating the relationship between goal setting and performance is self-efficacy, which is the belief in one's ability to successfully complete a task [8]. According to Lambie and Vaccaro [72], the more the student is interested in his or her current research, the more self-efficacy he or she gains. Particularly in PhD students, self-efficacy can be hampered by a fear of failure and/or by maladaptive perfectionism. Both often result in avoidance behavior in the face of academic adversities [57]: doctoral students with higher perfectionism or fear of failure try to avoid looking incapable by engaging in debilitating behaviors [109]. Furthermore, it was found that there is a significant relationship between perfectionism and academic procrastination [104]. Overall, more students than members of the general public state that procrastination of tasks is almost always an issue. Among students of different academic levels, graduate students were 3.5 times more likely to report the matter as a problem than undergraduates [92]. Goal-setting theory shows how the previously examined factors (i.e., research interest, self-efficacy, and self-worth) can moderate motivation, thus affecting commitment to goals and overall performance. From this

section, we drew the third requirement in the requirements table at the end of this chapter (Table 2.7).

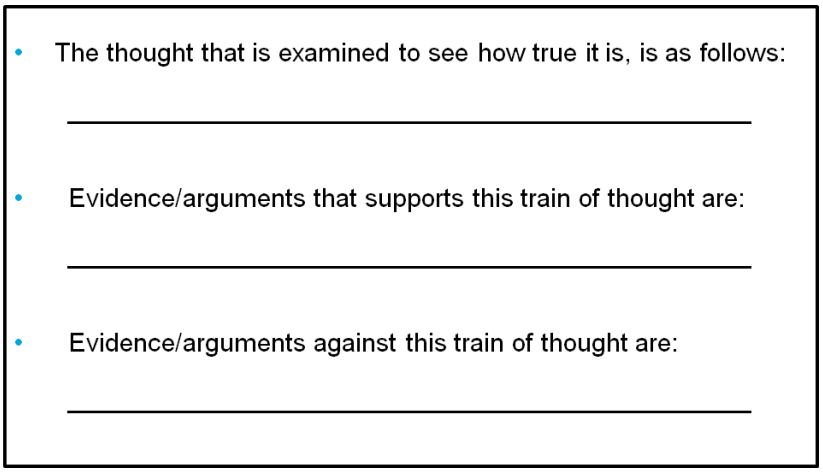


Figure 2.1: Example of an exercise for strengthening alternative thoughts, derived from [59].

2.1.2. Motivational barriers to doing the homework exercises

Strengthening thought exercises are usually completed either on paper (e.g., workbooks or papers handed to patients by psychologists) or in digital form by using a computer (e.g., in a word document) [59]. This is also the case for TU Delft PhD students, as stated by a PhD candidate's psychologist in the university. An example of such homework is shown in Figure 2.1.

The problems driving the lack of motivation to perform the exercises vary widely, and there is no concrete list of reasons behind such motivational issues [75]. Also, psychologists' opinions differ on which problems are relevant or most prevalent [11]. Therefore, we explored various sources to compile a list of nine potential motivational barriers. Table 2.1 shows a description of these barriers and identified strategies or theories to address them.

In this research, we narrowed our focus down to cover the first four motivational barriers (namely: disorganization, not convinced, low self-efficacy and perfectionism). This decision was made based on two reasons: (a) These four barriers were repeatedly mentioned in different papers, and (b) they align with PhD candidates' motivational problems described in [109]. These barriers and strategies to address them are used later in the focus groups to produce scenarios.

From this section, we came up with requirements 1 and 4, which are shown in Table 2.7 at the end of this chapter.

2.1.3. Value Stories

To investigate the problem more thoroughly, we focused on the stakeholders' values. By highlighting their personal values, we can take them into account to support them in the system design (Table 2.2). From the value stories, we derived requirements number 2, 6 (Table 2.7) and 11 (Appendix A.4).

2.2. Human Factors Knowledge

In this section, we examined three concepts related to our research: (a) PhD students as technology users, (b) human-chatbot interactions, and (c) how technology can motivate people. By covering these three points we have an understanding of how chatbot technology can influence the target group to perform actions.

Overall, PhD candidates differ from the average computer user in terms of technology usage. The technology usage gap includes several factors, among which are age and education gaps. Although the age gap of the average computer user is currently the most prevalent factor, it is expected to decrease as the generations shift [113] and to be surpassed by the education gap in the future. As for application usage, it was found that highly educated people tend to be better at using advanced technology for practical reasons, such as work and study, whereas people with less education tend to use simpler applications, mainly for entertainment, e-commerce, and communicating [113]. Furthermore, it was argued that the usage gap in education aligns with the knowledge gap [110], which states that people with less education perceive less knowledge from media

(e.g., television) than those with higher education [114].

For many, text-based communication is considered the default method of communication on the Internet; therefore, users will be accustomed to it when communicating with a conversational agent [62]. In their study comparing human–human interactions with human–chatbot interactions, Hill et al. [62] found that although humans sent more words per message to other humans, they sent more than double the amount of messages to chatbots. According to the authors, people show high interest in interacting with conversational agents, even though it is not as rich as human-to-human interaction. Furthermore, Brandtzaeg and Følstad [16] suggest that the motivations behind using a conversational agent will remain stable because they fulfill some human needs (e.g., providing assistance and social interaction).

Much research has been conducted on changing motivations toward technology usage. One method is through utilizing self-determination theory (SDT) (e.g., [108]). The main idea behind SDT is to examine how motivation can affect one's behavior. SDT includes three basic psychological needs that lead to increased motivation and performance when satisfied: competence (mastery of experience), autonomy (control of choices), and relatedness (feeling connected to others) [37]. The technology acceptance model, which is an extension of SDT, shows that people with higher autonomy and competence tend to be more motivated to continue using a technology in a work context [102]. Conversational agents in the mental health care domain have used various motivational strategies based on SDT. This is discussed in the next section.

From the three examined concepts above, we derived requirements 7 (Table 2.7), 9 (Appendix A.4).

Table 2.1: List of the gathered motivational barriers, their definitions, and identified theories or strategies to deal with each of them.

Motivational barrier	**Definition**	**Strategies or theories to handle it**	
Disorganization [11] [68] [61]	Some patients think that there are many tasks to do, and are not sure which one to start and when, to the point they could feel lost.	Checklist [11]	Start directly [68]
Not Convinced [11] [68] [61] [64]	Patients may refuse to do the exercises as they do not agree with the rationale. Some patients are not convinced that doing the exercises would gain them benefits in the long run.	Vicarious experience (social model) [9] [11] [68]	Studies and statistical arguments [109]
Low self-efficacy [11] [68] [44] [61]	Psychological distress can prevent patients from doing the exercises. They may lack confidence in themselves and think that they are not capable to do the homework. This can be specifically the case with higher symptom severity.	Raise self-efficacy [11]	Levelling [44] [68]
Perfectionism [11] [68] [61] [64]	Patients either do it all perfectly or do nothing. This mentality can hinder patients from performing the exercises as they lose too much time and may end up avoiding starting the exercises.	Simple reminder [11]	Behavioral experiment [11] [68]
Not understanding how to do the task [11] [68] [61]	Patients may perceive the tasks as being too difficult or too vague, thus being less likely to finish the assignments.	Scaffolding [2]	Levelling [11]
Forgetting the rationale [68] [11] [64]	Some patients tend to skip doing the assignment because they forget why they are doing it.	Record rationale next to the assignment [11] [68]	Establish a routine [11] [68]
Forgetting to do the tasks [68] [64]	Patients may simply forget to do the tasks. This could be due to different reasons, such as bad planning.	Associate the task with another task [68]	Establish a routine [68]
No enough time [11] [68]	Patients may think that they do not have enough time to do the homework. Their schedule is already busy, and they are not able to allot some time to complete the assignments.	Establish a routine [68]	Associate the task with another task [68] [11]
Impatient to get results [68] [74]	Some patients may feel that they should already get the positive outcome from therapy, as they have already done much before starting the exercises.	Gamify tasks [38]	Convince logically [68] [74]

Table 2.2: List of stakeholders and their value stories.

Direct Stakeholders	
PhD Candidates: The main users of the system. PhD students who show signs of depression symptoms.	- As a PhD candidate, I want to perform exercises to strengthen healthy thoughts, to improve my mental well-being, thus boosting my performance at work. - As a PhD candidate, I want the exercises to be understandable, so that I do not spend additional time on understanding. - As a PhD candidate, I want these exercises to be flexible, so that I can perform them whenever I like. - As a PhD candidate, I want my data to be handled privately, so that I can trust the treatment. - As a PhD candidate, I want a tool that can be used discreetly, so that when I do it at work, my colleagues do not necessarily find out. - As a PhD candidate, I want a tool that reduces therapist visits, so that I can have more time to work. - As a PhD candidate, I want to have a sense that the exercises will actually help improve my mood.
Indirect Stakeholders	
Therapists: PhD candidates' therapists, who are responsible for their mental improvement.	- As a therapist, I want a more flexible method to present thought-training exercises, so that they are more appealing to the candidates. - As a therapist, I want my patients to use a reliable tool to fill out the handouts, so that I can depend on it. - As a therapist, I want patients to do their homework because it forms the basis of sessions. - As a therapist, I want patients to understand how to do the exercise, so that they can complete it effectively. - As a therapist, I want patients to understand the potential benefit of the exercise, so they can have faith that it will help them. - As a therapist, I want patients to feel that they are able to complete the exercise, so lack of self-efficacy does not hinder them. - Overall, as a therapist, I want all my patients to have better mental well-being.
Supervisor/Project leader: Supervisors or project managers who are in charge of PhD students.	- As a supervisor/project leader, I want the PhD candidates to be in the best state of mind. - As a supervisor/project leader, I want the PhD students to perform their assignments correctly and within the time frame. - As a supervisor/project leader, I want the PhD candidates to feel comfortable around me.
Graduate School: The university to which PhD students and therapists belong.	- As a graduate school, I want all employees to be in the best state of mind, so that their productivity will be boosted. - As a graduate school, I want to create the best working environment possible.- As a graduate school, I want to reduce the number of dropouts.

2.3. Current Related Technology

Chatbot interventions have emerged from the recent improvement in conversational agents for mental health, to the point that some researchers have called them "the future of therapy" [46]. Many meta-analytical reviews have been conducted on conversational agents in the health care context [70] [88]. When investigating cCBT chatbots, researchers have reported higher overall adherence to treatment, which led to better overall psychological well-being (e.g., [82]). Two recent meta-analysis focused on conversational agents in mental health and found better treatment outcomes in terms of efficacy and acceptability compared to no treatment or information control [63] [55]). However, both papers agreed on the need for further research regarding the effectiveness of such technology in clinical settings.

In their literature review, Burger et al. [20] explored various e-mental health interventions that focused on depression symptoms. The authors reported that the majority of interventions (76%) were hosted on an online webpage, and 85% were based on CBT techniques. Furthermore, this software focused more on delivering treatment (about twice as much) rather than providing adherence strategies. The authors also argued that from 2000 to 2017, and despite the apparent increase in system numbers, technological advancements are lacking in the domain. In addition, they highlighted the need to conduct further clinical trials, especially in the case of more complex programs. To gain a better insight how CBT-based technology can operate, we examined some available computerized tools for mental health interventions. One well-known example of

a CBT-based chatbot is Woebot, which is a chatbot for depression symptoms in adolescents [51]. Woebot aims at delivering CBT through a series of short conversations. The chatbot does that by exposing the user with a concept related to CBT, such as cognitive triad or cognitive distortions. This is usually done through describing a situation where the concept can be applicable and then linking the user to a short video in which the concept is explained in more detail. Similar to Woebot, Wysa is a commercially available chatbot that is based on CBT practices and targeted toward young adults [65]. This chatbot tries to identify the emotions which the user expresses and reflect upon them using CBT and other self-help practices such as behavioral reinforcement. In addition, Wysa provides several therapeutic features, including dialectical behavior therapy, motivational interviewing, and mindfulness. It also contains tools that explain emotional self-support techniques. On the other hand, Help4mood follows a different approach by providing an embodied conversational agent in a program for mood self-monitoring [22]. Help4mood is targeted toward depression patients, and provides tools for self-monitoring moods, negative thoughts and behaviors. The virtual agent assists patients emotions-related questions to encourage them and promote user adherence. The program also provides summaries of user progress, which can give the therapist an idea about patients' progress [23]. Another program that follows a novel approach to deliver CBT is Sparx, which is a serious game targeted toward youth aimed at alleviating depression symptoms using CBT techniques [85]. In this game, the player tries to defy evil entities called GNATs, or "Gloomy Negative Automatic Thoughts", by overcoming obstacles in levels. The level contents are similar to the modules described before, where each level covers a concept of CBT. Players start each level by meeting a guide character, who sets expectations by providing the necessary psychoeducation before "fighting" the GNATs, or the negative thoughts (e.g., "you can't pass the exam"). At the end of each level, the guide explains how the new skills can be utilized to improve mood and fight negative thinking patterns.

Kretzschmar et al. [71] investigated some existing text-based cCBT systems, focusing mainly on Woebot, Wysa, and Joy [1]. The authors identified three minimum ethical standards that should be presented in a cCBT chatbot: privacy and confidentiality of data, efficacy of approach and safety (i.e., reducing the risk that a chatbot can induce to minimum). The authors also argued that the current automated systems are far from creating the experience of face-to-face conversations. They also claimed that the responses can be improved by providing experiences that are more tailored to users' needs. As far as we know, existing CBT chatbots focused on educating the user on the different mood-raising techniques and evidence-based concepts, rather than motivating patients (or doctoral students) to perform thought-strengthening exercises.

2.3.1. Motivation strategies used by chatbots

Motivating users to engage with a chatbot can be approached by utilizing different strategies. Woebot, for example, tries to facilitate a sense of accountability using different strategies, such as setting expectations for check-ins or sending personalized messages to the user from time to time, in an attempt to initiate a conversation [51]. Also, the chatbot uses emojis and animated images for encouragement to do the task or as a reward after finishing the task. At the beginning of each conversation, Woebot asks the user to indicate how he or she feels and then gives an appropriate empathetic response; the registered feelings are then presented in a chart where the user can track his or her mood change.

Chatbots in other domains applied similar methods to increase motivation. In the educational domain, gamification is often used to make the learning process more interesting [48]. For example, Benotti et al. [12] developed a chatbot for teaching the fundamental programming concepts to high school students. They gamified the learning process by letting the chatbots explain a murder case and showing suspect files and asking the students to implement a counter chatbot that could answer the questions regarding the case. Then, final points were calculated based on how many correct answers were given by the chatbot. Another chatbot focused on using the nudging theory to nudge students into reinforcement behaviors to solve procrastination issues [103]. An example from the health-promoting domain is described in [47]. In order to motivate users to change, the authors integrated the efficiency model of support, which investigates five causes for not benefiting from intervention systems: usability, engagement, fit, knowledge and implementation [106]. Therefore, they approached engagement in their developed chatbot by integrating motivational strategies such as immediate feedback, activity tracking, encouragement of certain activities done by the user and user mood tracking. To sum up, Table 2.3 provides a glimpse of how various motivational strategies can be used in a chatbot to motivate a user to act, based on SDT. The table was built upon Cook and Artino Jr [32] work, where the authors presented a model relating motivation to SDT. In addition, we expanded the model with additional motivational strategies that are used in computerized cognitive behavioral therapy, as well as the systems that use these strategies.

Table 2.3: Motivational strategies that can be used in conversational agents based on SDT, built upon [32].

Self-determination theory		
Competence	**Autonomy**	**Relatedness**
• Feedback [22] [85] • Rewards [51] [65] [85] • Goal setting [51] • Self-monitoring [51] [22] • Self-help practices [65] [22] • Gamification [85]	• Sense of control [65] [85] [51] • Rational or education [51] [65] [22] • Verbal Persuasion [51] • Tailoring to individuals [51] [65] • Nudging [103]	• Anonymized community [65] • Social Modelling [85] [51] [65] • Assuring confidentiality [65] [51] [22] • Showing empathy [51] [65] • Verbal encouragement [51] [65]

From the current related technology, we derived requirements number 5 (Table 2.7), 11, 12, 13 and 21 (Appendix A.4).

2.4. Exploration

The aim of this section is to collect requirements that can be presented in the system. To achieve this, we followed a scenario-based approach that included two focus groups: the experts and PhD candidates. We did a focus group with domain experts (i.e., psychologists) at the beginning to grasp a comprehensive understanding of how to deal with PhD students' possible motivational barriers, which was the main focus of this group. Next, we did the focus group with PhD students, which aimed to understand their acceptance of such a system. Since we did not have a running prototype at that stage, we developed mock-ups (Appendix A.2) in which we could visualize design claims. Furthermore, we developed four personas and six scenarios, where each scenario covered a situation in which a chatbot can be used. These scenarios (alongside one persona) were used in the focus groups to acquire their thoughts and opinions. The results obtained from the focus groups were used to refine the final list of requirements, which is the outcome of this chapter.

2.4.1. Focus groups

In order to gather opinions on various strategies to develop the final requirement list, we met with 2 focus groups: psychologists and PhD candidates. Before meeting with the focus groups, we prepared six scenarios, four of which covered the motivational barriers that we established in section 2.1.2 (namely: disorganization, not convinced, low self-efficacy, and perfectionism). The other two addressed technical barriers to using the system (platform and interactivity). Each scenario represented a problem, and we produced two contrasting strategies that the chatbot can use to address each problem. In the focus groups, we asked about the participants' thoughts and preferences regarding those strategies. Figure 2.2 shows a summary of how the scenarios were distributed between the focus groups.

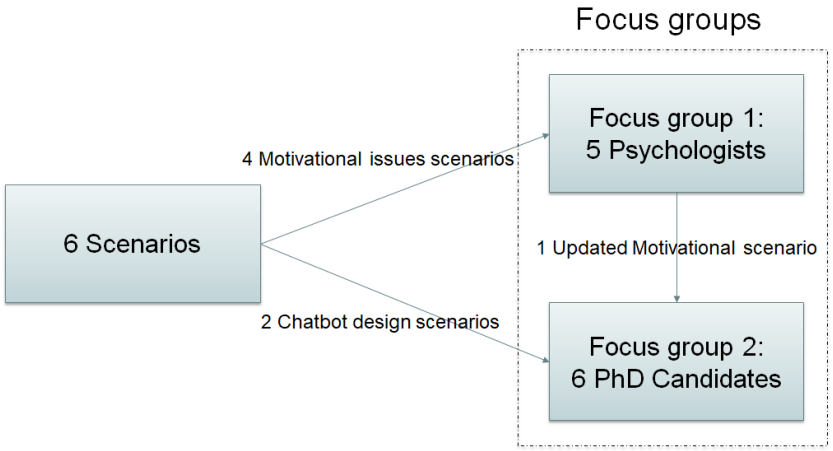


Figure 2.2: Diagram of how scenarios were used in the focus groups.

For the first focus group, we wanted to meet with five clinical psychologists, one of whom is specifically treats PhD students. Due to time constraints on the side of the psychologists, the focus group took the form of an online questionnaire that included the four motivational barriers described above and required about 17 minutes to complete. For each scenario, one brief video with a voice-over was prepared. The video explained (a) the motivational barrier in a story format and (b) two strategies for how a chatbot could deal with such an issue. The psychologists were then asked to write down their thoughts and preferences regarding the suggested strategies. After collecting all the replies, we pooled the answers and derived arguments supporting and opposing the strategies. These were used later to reflect upon the scenarios.

For the second focus group, we gathered six PhD candidates from different TU Delft faculties for a session that lasted about an hour. We began the session by showing an updated scenario and strategies using the psychologists' arguments (the "not convinced" scenario), followed by two scenarios about chatbot design. We asked the participants about their thoughts and preferences regarding the three scenarios. The session also included a general discussion of two main points: (a) motivational barriers that could prevent PhD candidates from performing thought-strengthening exercises and (b) technical barriers to using such a chatbot for that purpose.

The described focus groups were approved by TU Delft Human Research Ethics Committee on 19-07-2019, with the application ID number 857.

Table 2.4: A description of a persona who is Bill.

Demographics	
Name	Bill
Gender	Male
Age	30
Department	Computer science
Years since the program started	2
Relationship status	Married
Number of children	0
Motivation	
Hobbies	Playing musical instruments (specifically, piano and bass)
University	Ambitious and has a good academic status, which he struggles to keep. He is not very interested in his current research.
Personal challenges	Bill thinks his supervisor's demands are higher than his capabilities, but he tries to cope with the work stress. He always feels down when he gets negative feedback on submitted tasks. The main reason behind that is last-minute submissions for the tasks; he does not have enough time to adhere to the required quality.
Social Environment	
Background	Bill is from a middle-class family. He is an international PhD candidate doing research in the Computer Science Department. He lives alone in an apartment near the university. His family and wife live in the US.
Supervisor & Colleagues	Bill has casual relationships with his colleagues and supervisor. From time to time, he joins his colleagues on Thursdays evenings for a beer. His supervisor is available most of the time to answer his questions and provide him feedback.

2.4.2. Scenarios and results

In this section, we start by introducing a PhD student persona, which was used to construct the scenarios. Then, one scenario and its strategies are described as an example, alongside a summary of the experts' and PhD students' opinions on each strategy. The other scenarios details can be found in Appendix A.3. After that, the final suggestions based on the focus groups' arguments for all six scenarios are discussed. At the end, we

present a general discussion about the focus groups.

Persona
This persona of a PhD candidate (Table 2.4) was used in the next section to construct scenarios related to our research. The persona is based on a review paper by Sverdlik et al. [109], who investigated several factors that could affect the mental health of PhD students (see Appendix A.1 for three other personas based on the paper).

Scenarios structure
Next, we describe one scenario, alongside the focus group opinions. Each scenario includes a problem in the form of a story of our persona Bill, two strategies to address the problem, the focus groups' answers regarding each strategy, and our final suggestions from the argumentations. Table 2.5 shows the aforementioned process for one scenario (disorganization). The other scenario descriptions and argumentations can be found in Appendix A.3.

Table 2.5: An example of a scenario and the arguments derived from focus groups. [Psychologists]

Scenario 1: Disorganization	
Bill (i.e., the PhD student persona) had already met with a therapist who convinced him to change his negative thoughts and had decided upon alternative thoughts. The therapist then gave him exercises to train their alternative thoughts, but he felt lost when he attempted to do the exercises. He thinks there are too many things to do, and he does not know where to start.	
Strategy A: Checklist	
Description	**Arguments supporting A:**
- The chatbot suggests a weekly checklist of Bill's tasks. - Bill can add, modify, or delete tasks on the checklist. - Bill can access and complete any task on the checklist. - When Bill completes a task, it is marked off in the checklist.	- Control and choice can provide more motivation. - The client will not be bothered to do the tasks in a specific order. - The list provides an overview of what is left.
	Arguments opposing A:
	- It requires decision-making every time. - It could enhance avoidance behavior. - The client may not be able to process the structure, as his working memory is less available.
Strategy B: Start directly	
Description	**Arguments supporting B:**
- The chatbot asks Bill to complete a specific task. - The chatbot introduces a new task once Bill completes the current one. - Bill can only access the given task unless he asks to do a different one.	- No decision is required, so there is no extra burden. - Focusing on only one task at a time may be better when the client is unmotivated. - This strategy helps to overcome avoidance out of fear.
	Arguments opposing B:
	- This option offers the client little control. - The client may think that the tasks appear suddenly.
Suggestion	
Combine both strategies and make B the default strategy. However, Bill can have the option to change to scenario A whenever he likes (and vice versa).	

Results
The psychologist's arguments regarding the various issues provided better insight into the problem and how to approach creating a solution. It also aided in understanding how to address a motivational barrier (requirement number 4 in Table 2.7). Table 2.6 summarizes suggestions for all scenarios, which were concluded from psychologists' arguments. The scenario definitions can be found in Appendix A.3.

Table 2.6: Suggestions made for each scenario, based on focus groups argumentations.

Scenario	Suggestions
Scenario 1: Disorganization	Because providing a sense of control is important in this case, Bill can toggle between two methods: whether to let the chatbot suggest one task at a time, or ask the chatbot to provide a checklist of the remaining tasks. The first method can be the default option.
Scenario 2: Not convinced	Attempt to convince Bill using studies and statistics, as it generally works better for PhD candidates. On the other hand, finding a social model that fits all students could prove difficult.
Scenario 3: Low self-efficacy	Although providing verbal encouragement and exercises for raising self-efficacy could be less effective with severe cases, it provides a safer solution in the long run. However, the activities should be more personalized (e.g., "List 3 things you did well last week").
Scenario 4: Perfectionism	Remind Bill to not seek perfection in his answers. In the case the problem persists, the chatbot can ask Bill to do a behavioral experiment, in which he needs to deliberately make mistakes in the task and finish it within a certain time frame.
Scenario 5: Interactions	Most communication should be done through closed-ended questions. However, for some interactions, there can be an option for Bill to elaborate, after which the chatbot can confirm that it understood Bill's intentions.
Scenario 6: Platform	Hosting the chatbot on a secure website can make it more accessible to the students, but keeping in mind that privacy policies should be transparent (e.g., how the data is handled, where it is stored, etc.)

The PhD candidates agreed with the expert feedback on the scenarios. They also argued that in the case of presenting a social model, it does not need to match in terms of demographics but in story so that a PhD student can identify with it. If not, this may lead to an opposite effect of demotivating. As can be seen in Scenario 5 in Table 2.6, the PhD candidates preferred closed-ended questions with the chatbot due to it being faster and more intuitive (requirement 5). However, they also suggested that it could be beneficial to add an option that allows the student to elaborate when faced with multiple choices, and the chatbot can then confirm whether the intention was understood or not (requirement 15). Another detail about the system implementation was regarding the platform used. One presented idea was to implement the system as a mobile application, but this may prove difficult to use due to the nature of the thought-strengthening assignments. Therefore, hosting the chatbot on a website that is also mobile friendly can make it more accessible, especially for those who want to try it at the beginning (requirement 8). This sparked a discussion about the privacy of the chatbot, as some sensitive data could be communicated using the system. Some of this sensitive information is needed to construct negative and positive thoughts. An example of that was having negative thoughts toward their supervisor or project leader. Revealing other data (such as gender) may lead to narrowing down the identity of the negative thought possessor. Therefore, the system should be explicit about how the data is handled and where it is stored (requirement 6). This aligns with Kretzschmar et al. [71] privacy concerns regarding virtual agents. Furthermore, PhD candidates argued that some users may not believe the chatbot

can help them, as it seems to be a solution for “simple issues.” Therefore, it is important to set expectations at the beginning regarding what they can expect to gain from using the chatbot. For example, this could mean mentioning that the chatbot cannot replace rich face-to-face interactions with a psychologist, but rather supports it Kretzschmar et al. [71]. Moreover, they preferred having the chatbot as a coach to support them doing the exercise, rather than a peer. This is due it being known that it is actually a bot, not a real person, which can be harder to identify with (requirement 11).

The candidates also suggested providing an incentive for doing the exercise again. This includes strategies such as a reward system (requirements 11, 20) and progress monitoring (requirement 16). Other suggestions related to interactivity were integrating images and “emojis” into the conversations to facilitate better communication (requirement 17) (alligning with [25]), and being able to tell jokes; however, the humor should be balanced with seriousness [26] (requirement 19). Also, as the shown mock-ups were early concepts, the participants commented on the “form-like” prototype and preferred a more visually appealing design that can include a customization option (requirements 13 and 21).

2.5. Requirements

Based on examining the relevant background, literature reviews and focus group meetings, we compiled the final list of requirements. The requirements were divided into major and minor requirements, where the major ones (requirements 1-8) were considered to essential to answering the research questions (shown in Table 2.7). On the other hand, the minor requirements (listed in Appendix A.4) were the desirable functionalities, though not vital. In this research, we focused mainly on fulfilling the major requirements.

Table 2.7: List of the systems’ major requirements.

1. Aid PhD candidates in doing thought-strengthening exercises.
2. Include a means to perform the thought-strengthening exercises.
3. Identify PhD students’ motivational barriers that could obstruct them from doing thought-strengthening exercises.
4. Address PhD students’ motivational barriers.
5. Understand the PhD students’ inputs, and provide proper replies.
6. Be transparent about the privacy policies.
7. Provide the necessary psychoeducation to the PhD candidates in order to be able to complete thought-strengthening exercises.
8. Be easy to access.

3

Design

The third chapter's aim is to answer the third research question:

*What **design** would meet the requirements?*

We start this chapter by presenting the suggested solution from the user's perspective. This is done by showing how the proposed software can fulfill each requirement. After that, we explore models of homework exercises used by psychologists in real sessions and show how these models were transferred as a session flow in the system. Finally, we discuss the content of the session flow and link it to relevant literature. The output of this chapter is a proposed system, which we can base our prototype on.

3.1. Proposed solution

In this section, we show our solution for fulfilling the system's requirements. For each requirement, we describe how our solution works and looks and how it would lead to meeting the requirement. More information regarding design recommendations (proposed by Montenegro et al. [88]), comparisons with other chatbots, and system categorization can be found in Appendix B.

Requirement 1: Aid students in doing thought-strengthening exercises

The proposed solution is to motivate the PhD candidates to complete thought-strengthening exercises by using a chatbot to help overcome the candidates' motivational barriers to complete the exercise. Chatbots' usefulness in improving motivation has been established in the previous chapter. Figure 3.1 shows the system's interface, which is separated into two windows: the intervention window (left) and the support window (right). The idea behind the intervention window is to show the current exercise that the candidate needs to complete, while the support window includes interactions with the chatbot. The decision to have two separate sections allows the student to access and interact with the chatbot and complete the exercise at the same time. In addition, it allows the student to access the log of his or her interactions with the chatbot. One concern regarding this approach is the possibility of increasing mode errors [91] — that is, attempting to perform one window's functionality in the other window (e.g., writing an exercise solution in the chatbot window). This type of error is reduced by showing a clear distinction between both windows: the support window looks like a conversational interface, and the intervention window appears as a form (homework) that needs to be completed [18] [107]. A similar solution was also used in various other agent-related systems (e.g., [13] and [96]).

Requirement 2: Include a means to perform the exercises

The PhD candidates can complete the thought-strengthening exercises through the intervention window on the left (Figure 3.1). This window shows the current exercise questions and asks the student to answer them, where the answers can be registered and accessed later by the student. When the current exercise is completed, the next one is made available.

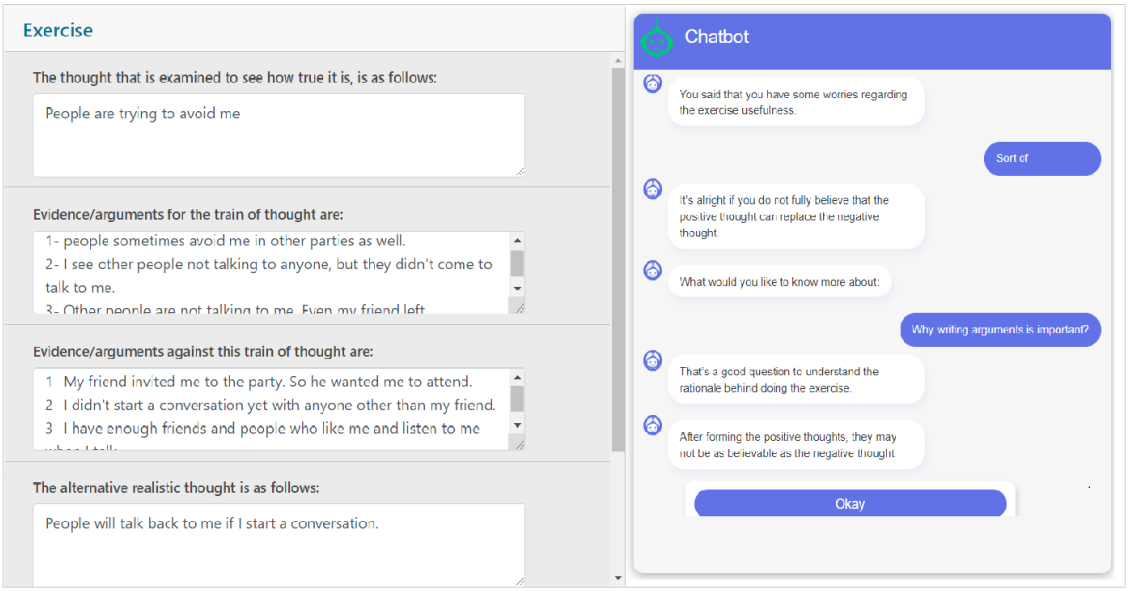


Figure 3.1: An overall look at the proposed solution, with the intervention window (i.e., the exercise) shown on the left and the support (i.e., the chatbot) on the right.

Requirement 3: Identify motivational barriers
The reason behind the student's lack of motivation to complete the exercise should be identified before tackling it with appropriate strategies [68], so that the candidate and the chatbot can identify the applicable barrier together. This will be done by the chatbot asking the student to choose the most applicable reason for him or her. For example, Figure 3.2 shows how the chatbot confirms the motivational barrier with the candidate (in this case, the exercise's rationale).

Requirement 4: Address motivational barriers
After the motivational barrier is identified, the chatbot will address the barrier using the strategies explained in the previous chapters. In the example in Figure 3.2, the empathetic response was that "It is normal to not fully believe that the exercises are useful." To address the barrier, the chatbot will attempt to convince the student of the exercise's usefulness and the rationale behind it by determining what may convince him or her (e.g., evidence from studies or arguments that support its importance) and then show the arguments and interacts with the student based on the strategy.

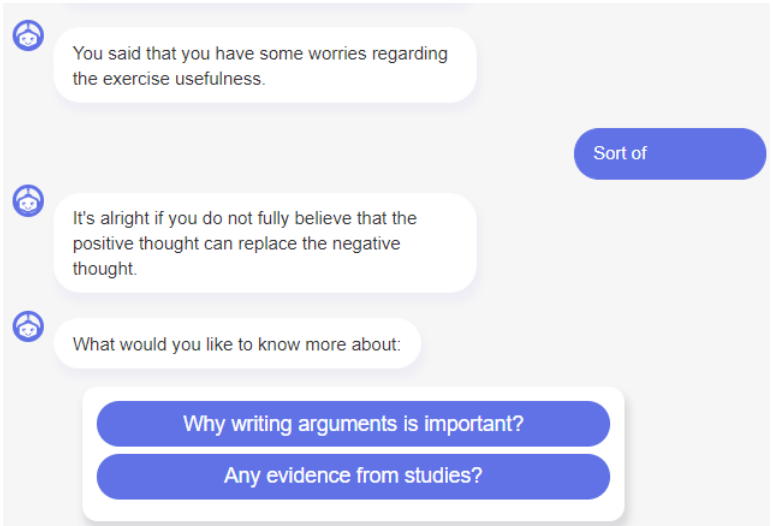


Figure 3.2: How the chatbot deals with a motivational barrier. In this case, the barrier is that the candidate is not convinced of the rationale behind doing the exercise.

Requirement 5: Understand the inputs, and provide proper replies
As Figure 3.2 shows, the interactions mainly take place through closed-ended questions to avoid misunderstandings [87]. For each question or input request, the chatbot provides the student with two (or often three) choices from which to choose. The chatbot navigates through its dialogue trees based on the user's choices. In some cases, the chatbot will ask the user to type an answer (e.g., to enter a name or answer questions) to raise the student's self-efficacy.

Requirement 6: Provide the necessary psychoeducation
During the first session with the agent, the chatbot will start by explaining the exercises and why they should be done (e.g., Figure 3.3). This can help in giving context to the rationale behind the exercises and the chatbot's role in the intervention. Another important point is for the chatbot to provide the necessary information to complete the exercises. This is done at the beginning of each new exercise.

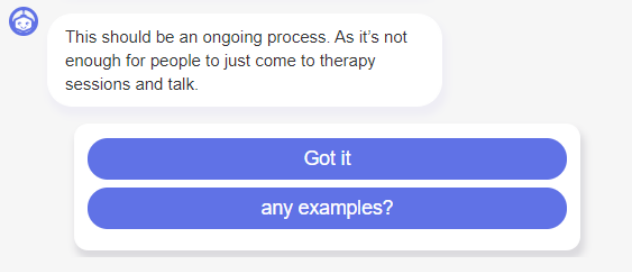


Figure 3.3: The chatbot provides psychoeducation to the candidates about the exercises' rationale.

Requirement 7: Be transparent about the privacy policies
The exercises and interactions with the chatbot may include sensitive data that the PhD students would prefer to keep confidential, so they might hesitate to complete the exercise. Therefore, during the first session, the chatbot will clarify where the data will be registered and who can access the data, and it will ensure users of confidentiality. In a later section, we discussed how the chatbot assures confidentiality.

Requirement 8: Be easy to access
As accessibility is important for PhD students, the system can be accessed by using a Web browser rather than a standalone application. This gives the users a chance to try it without the need to set up the system. This aligns with Foggs' model [53], where the easier a task is, the more likely the action will be taken. Therefore, the website displays only the two windows shown in Figure 3.1 without any additional distractions.

3.2. Exercise sessions

In the previous section, we provided guiding principles to show how the proposed solution can address the requirements. In this section, we started by linking the real-life exercise sessions with the exercise done through the proposed system, then discussing how they were integrated in the actual design of the system.

In their book about using homework in CBT, Kazantzis et al. [68] identified four primary stages of a guiding model for a homework session: reviewing the last assignment, designing the current assignment, assigning the assignment, and finally completing it. Reviewing the assignment includes a discussion of the past assignment between the psychologist and the patient, which is out of this project's scope. The second stage is designing the exercise, i.e., deciding on the type of homework to be completed by the patient and providing information for the user to understand its rationale. After that is the assigning stage, which requires a seamless transition from the last stage as the two stages sometimes overlap. Assigning can include preparing the patient to complete homework, providing the necessary psychoeducation to help the patient to do the homework, or adjusting the homework based on the patient's status or preferences. Last is the homework-completion phase, during which the patient completes the homework. The last three stages (i.e., designing, assigning, and completing) were adopted to our system and presented in every new session for each homework exercise.

A session flow that includes the three previously mentioned stages can be seen in Figure 3.4. The session flow explains how a typical interaction session between the proposed system and the PhD candidate would go. This flow was inspired by the interactions for Woebot [51] and Wysa [65], and it was further integrated with the three stages from the guiding model for homework.

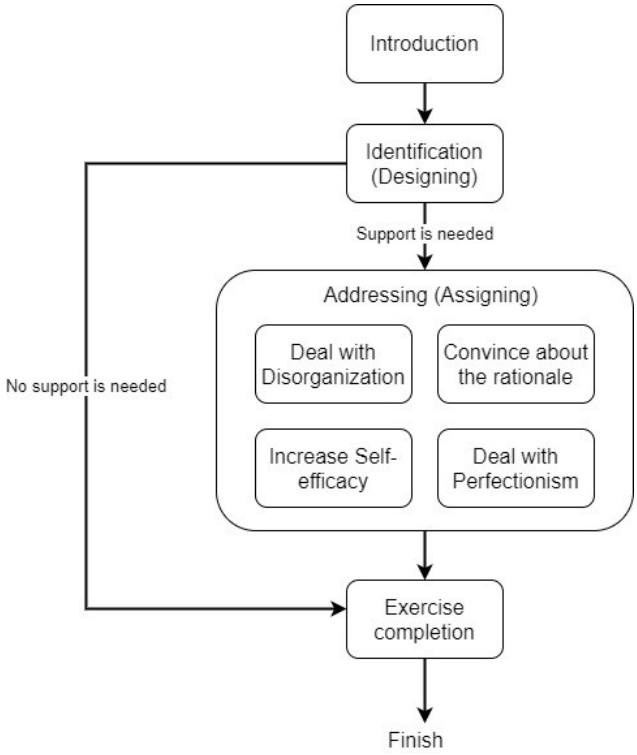


Figure 3.4: A typical session interaction flow that the chatbot uses in every session. The words between brackets refer to the stages in the guiding model, as defined in Kazantzis et al. [68].

3.2.1. Introduction

The first step during the interaction is the introduction. At the beginning of the initial session, the conversational agent will start by greeting the user and then asking for a name to call the user. The same name will be used later in the conversation to build bonds between the user and the agent [26]. To familiarize the users with the system [49], the chatbot will introduce itself and what it can offer in the process of completing thought-strengthening exercises. After that, the chatbot will ask if the user is familiar with the term "thought-strengthening exercises," which it will or will not explain based on the user's choice. The term's description, as taken from [59], will act as a refresher to the student, rather than it explaining a new concept, as it is assumed to have been explained by a psychologist beforehand. Furthermore, ethical and legal considerations are made regarding the user's privacy to ensure confidentiality of the registered data. We started by looking at the ethical manifesto for software in e-health [89] and the EU General Data Protection Regulation (GDPR) [116]. The conclusion was that the chatbot will inform the candidate about where the data will be stored and who can access the data before starting the exercise.

During later sessions, the introductory sequence will be briefer than the first session because it will be a preface to a new exercise session, rather than an overall introduction to the homework exercises. Therefore, the chatbot will start with a simple greeting followed by asking if the user is ready to start the exercises. Figure 3.5 summarizes the sequences of the introduction in the two different cases.

The identity chosen for the chatbot was a simple avatar of a smiling robot that acts friendly to the user through text. This decision was made for two reasons: to avoid possible gender-related influences on the human–agent interactions [76] [60] and to avoid uncanny valley effects that may occur during interactions with humanoid avatars [31].

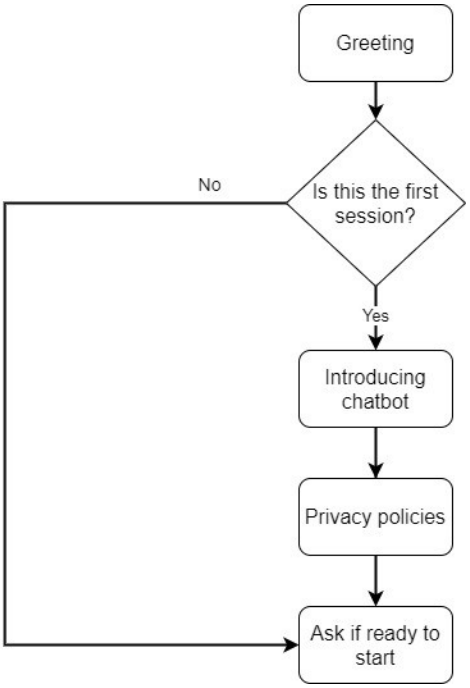


Figure 3.5: How the chatbot starts and introduces each session.

3.2.2. Identification (designing)

The next phase is the identification phase, which corresponds to the designing stage in the guiding model. This phase consists of three stages: (1) psychoeducation on the current exercise, (2) determining if the candidate needs support to complete the exercise, and (3) determining the candidate's motivational barriers toward completing the exercise. In the first stage, the chatbot will explain the exercise that the student must complete. The explanations of the exercises were derived from self-help books for applying cognitive therapy techniques [59] [101]. The explanation will differ by exercise based on the exercise's nature. Usually, it will consist of a description or example as well as tips for completing the exercise. The aim of the second stage is to check if the candidate needs additional support toward his or her motivational barriers. This can be done by directly asking the candidate to rate how likely he or she will be to complete the exercise (from 0% to 100%) [11]. If the user rates the likability higher than 90%, then the chatbot will ask to complete the exercise directly, without the need for identifying and addressing motivational barriers. Otherwise, the agent will proceed with the third stage, i.e., determining a motivational barrier. In this stage, the chatbot will ask the student to choose one of the four motivational barriers we identified in the previous chapter. The chosen barrier will then be addressed using one of the motivational strategies. The sequence of the three explained stages is shown in Figure 3.6.

3.2.3. Addressing (assigning)

After the candidate's motivational barrier is identified, the next step for the chatbot is to address this barrier. The chatbot would first acknowledge the barrier then show empathy toward it, as expressing empathy in conversational agents context is preferred over the impassive provision of advice [78] [111]. In our running example, the empathetic response that the chatbot delivers is "It's alright if you do not fully believe that the positive thought can replace the negative thought." Empathetic responses also are expressed for different situations (such as asking a question or requesting additional information from the chatbot).

Next, the chatbot addresses the concern barrier. We already examined motivational strategies to use for the barriers in the previous chapter; therefore, we are only going to discuss how the chatbot can integrate the strategies. The strategies used by the chatbot can be divided into two categories: information-based and exercise-based. In the information-based strategies, the chatbot tries to interact with the candidate to confront the barrier before doing the exercise, without changing the exercise's behavior or appearance. In contrast, exercise-based strategies change the exercise behavior. These changes are explained by the chatbot beforehand to clarify why the adjustments will be made. Table 3.1 summarizes the categorization of the strategies used by the chatbot. Additionally, Figure 3.7 provides a summary of the flow of addressing barriers.

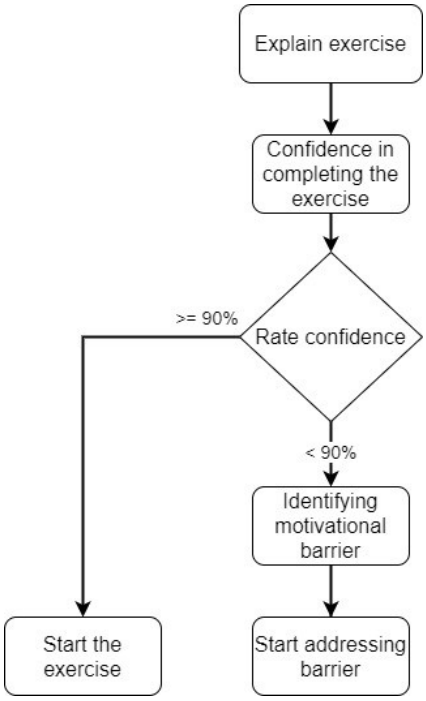


Figure 3.6: Sequence for rating the likelihood of completing an exercise and identifying the user's motivational barrier

Table 3.1: Motivational strategy and how they can be used by the chatbot.

Category	Barrier	Strategy
Information based	**Not convinced**	As PhD students prefer showing studies and statistics to establish the usefulness of an exercise and the rationale behind it, the chatbot shows relevant statistics from studies (e.g., [68], [75], [56]).
	Low-self-efficacy	Based on Bandura's self-efficacy model [9], the chatbot attempts to raise the candidate's self-efficacy through verbal encouragement and persuasion and vicarious experience.
	Perfectionism strategy A	The chatbot provides a simple reminder not to seek perfectionism in doing the exercise, taken from [11]. If this strategy does not work, the chatbot will implement the Perfectionism B strategy instead.
Exercise based	**Perfectionism strategy B**	The chatbot tells the candidate to do a behavioral experiment and adds a time limit to the exercise, then asks the candidate to complete it. This is followed by a small discussion with the chatbot regarding whether the candidate is satisfied with his or her answers.
	Disorganization	Rather than showing the whole exercise in the intervention window, only one question at a time will be shown and will need to be answered before the candidate can proceed to the next one

3.2.4. Exercise completion

At the end of the interaction, the chatbot will ask the PhD candidate to start doing the exercise. As explained in the previous section, the exercises are completed using the intervention window, which is shown on the left. To gather the homework exercises, we looked at two self-help books on applying CBT [59] [101]. The homework questions and texts were taken from the two books and given a similar format (e.g., a text field for questions). After completing the exercise, the PhD candidate can submit the exercise.

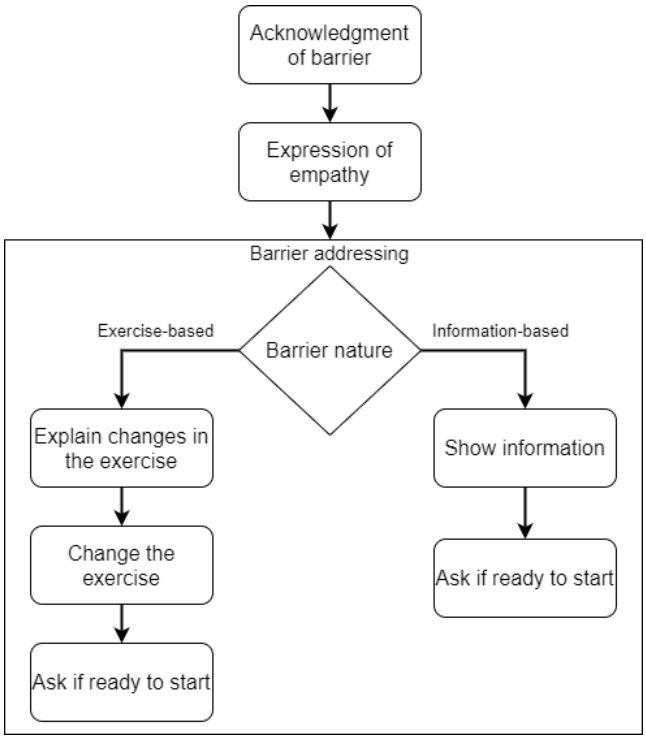


Figure 3.7: Session sequence for addressing motivational barriers.

4

Evaluation

To conduct an experiment on the developed prototype, we followed Field and Hole [50] proposed approach for designing and reporting an experiment. Therefore, we will start by explaining and arguing the formulated hypotheses, which includes an explanation of the developed prototype to use in the experiment. After that, the methods used in the evaluation are described. This is followed by describing the results obtained from the analysis and exploration of the data. The chapter ends with a discussion section on the results. Overall, this chapter tries to answer the third research question, which was the following:

*Can such a conversational agent increase doctoral students' **motivation** toward thought-strengthening exercises?*

The research question defined before was made more specific, and the hypotheses were defined accordingly to cover two aspects: perceived usefulness and self-efficacy regarding the thought-strengthening exercises. We decided on these factors because the examined motivational barriers are relevant to one or both of the factors. Various models have established a link between the two factors (i.e., perceived usefulness and self-efficacy) and motivation. According to the Intrinsic Motivation Inventory [83], higher perceived usefulness can lead to higher overall motivation. As for self-efficacy, Bandura [5] argued that an important component that affects motivation is self-efficacy beliefs. Moreover, perceived usefulness and self-efficacy are two of the variables of the behavioral intention to use, according to the technology acceptance model (TAM), the first being an original variable and the latter an external one [35]. The Health Belief Model also showed a primary role of self-efficacy in increasing the likelihood of engaging in health promoting activities [29]. In addition, Baker et al. [3] proposed strategies to aid research in the context of eHealth. One of these strategies is to use short-term (proximal) outcome measures to evaluate the system and more quickly and efficiently.

Strategies for addressing motivational barriers regarding thought-strengthening exercises were shown to help clients overcome barriers [68]. Furthermore, the strategies have been used and explained in various CBT-related sources (e.g., [59] and [11]). On the other hand, text-based conversational agents comprise two primary aspects: a content side (which, in this case, can be the textual content of the strategies to address barriers) and a form side (which is the interactivity of the presentation format) [27]. As established in previous chapters, conversational agent supports has more potential to improve humans' motivation compared to textual supports. Therefore, the hypotheses were formulated as follows:

H1: *The users' perceived usefulness of the thought-strengthening exercises is highest in the conversational agent support case, followed by the text support case and then the no-support case.*
H2: *The users' self-efficacy toward the thought-strengthening exercises is highest in the conversational agent support case, followed by the text support case and then the no-support case.*

Prior to starting the experiment, we conducted a pilot run with 20 workers with that setup to check if it worked properly, and we fixed the issues that arose. The 20 workers were paid, but their data were not included in the analysis.

Ethical approval for conducting the experiment was granted by the TU Delft Human Research Ethics

Committee on 28-01-2020, with the application ID number 1002. Additionally, the evaluation was registered through the Open Science Framework (OSF) registration platform before data collection began [94].

4.1. Methods

This section explains the methods used for evaluating the hypotheses. First, we present an overview of the experiment. This is followed by the participants, materials and measures description. In addition, the developed prototype and decisions for utilizing it in the experiment are discussed. Finally, the data analysis strategies used for obtaining the results are defined.

4.1.1. Experimental design

The experiment setup had a double-blind mixed design and included three conditions: the no-support group, who received the intervention only (i.e., the exercises to strengthen a new thought); the text-support group, who received the intervention and support strategies to address motivational barriers through text (i.e., they were not delivered by a chatbot); and the chatbot-support group, who received the intervention and the same support strategies but delivered by a chatbot. These three conditions enabled separation of the content of the motivation (i.e., with the no-support case) and the presentation format (in the case of text support). The participants were randomly assigned by the computer to one of the three equally sized groups. Additionally, they were assigned to a scenario with which to perform the exercise. The participants of each group were asked to fill out pre- and post-experiment surveys on the measures (namely, perceived usefulness and self-efficacy) that was used to compare the three groups.

4.1.2. Materials

Prototype

The prototype design was based on the design description in the previous chapter. To evaluate the hypotheses, however, we redefined some aspects of the system implementation. The first and most noticeable change concerns the intervention and support windows. In the prototype, the user had to first interact with the chatbot (Figure 4.1), which is the support that includes the motivational strategies, and then do the exercise (i.e., the intervention in Figure 4.2). This decision was made to ensure that people were exposed to the chatbot before doing the exercise (i.e., interacting with the chatbot is a system initiative). In the original design, the exercise and the chatbot are shown at the same time (the interaction is user initiative). In addition, we only decided to include the three barriers that have an information-based strategy, because the exercise-based strategies will change the exercise, thus changing the intervention, which should be intact in the three conditions. In both conditions (i.e., the agent and the text), the highest-rated barrier was addressed. Three motivational strategies were presented to address each barrier, which included acknowledgement and empathy. The first strategy is the strategy defined in the previous chapter (based on verbal persuasion). The other two strategies (based on authority and the social model) were taken from Cialdini's principles of persuasion [30]. To implement the chatbot, Rasa NLP, which is an open source natural language processing program for matching intents with input, was used. More technical details regarding the chatbot implementation (including the architecture and a dialogue tree) can be found in the Appendix C.1.

Motivational strategies

As stated before, the motivational barriers were only tackled in the chatbot- and text-support conditions. The methods of tackling each barrier were argumentation, persuasion based on authority, and persuasion based on a social model. The arguments for the three barriers were taken from a self-help book by Greenberger and Padesky [59] and a textbook on CBT, targeted toward psychology students, by Applebee and Langer [2]. The authority text was taken from the same self-help book in addition to Kazantzis et al.'s textbook on using homework with CBT [68]. In addition, the social model texts for the different barriers were shortened versions of stories that apply to the given motivational barrier, which were taken from the second textbook.

Other materials

Before starting the experiment, the participants were asked to agree on a consent form that explained the study and how the gathered data would be used (see Appendix C.2).

The experiment included two scenarios representing different situations that could provoke negative and positive thoughts. Each scenario included three negative and three positive thoughts, which the participants had to rate. The two scenarios were included to generalize across scenarios by avoiding effects that might

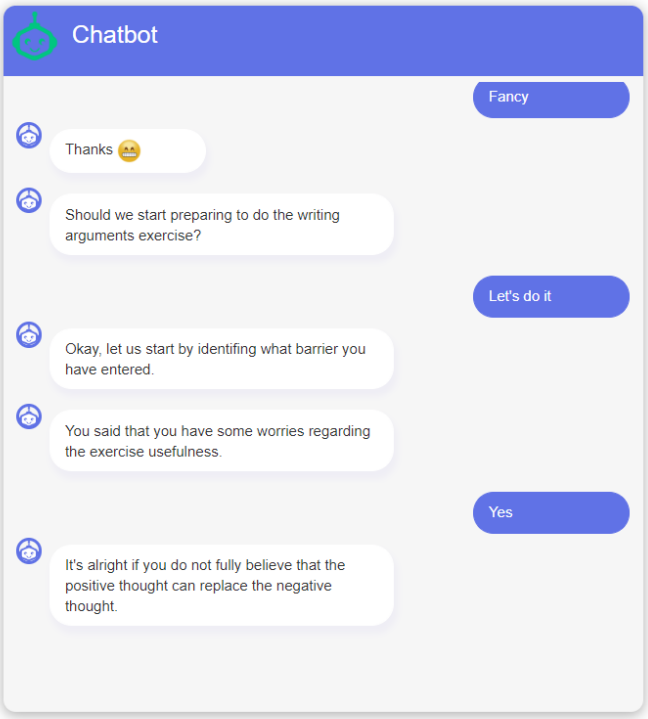


Figure 4.1: Screenshot of the interaction with the chatbot. The motivational barrier tackled in this case is if the participant is not convinced of the exercise's rationale.

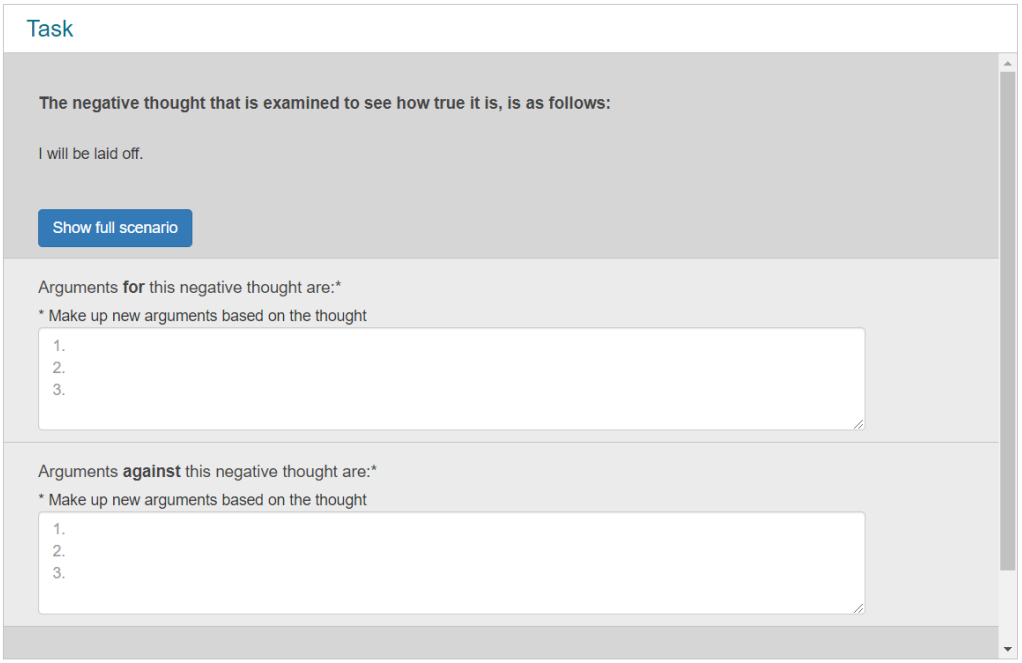


Figure 4.2: The exercise shown for the three conditions. In this case, the scenario can be displayed using the "show full scenario" button. Also, the highest rated negative thought for the participant can be seen at the beginning.

be linked to a specific scenario, thus increasing the material's external validity. Participants were randomly assigned to one of the two scenarios to avoid systematic biases caused by differences between the two scenarios. The scenarios, along with the thoughts, were taken from the aforementioned self-help book Greenberger and Padesky [59].

To simplify the conceptions, the exercise was called a "writing arguments exercise" instead of a "thought-strengthening exercise." That name was chosen because the participants wrote arguments for and against the thought, which is the first exercise for strengthening alternative thoughts in Greenberger and Padesky [59]. In this exercise, the participants were asked to provide evidence (i.e., arguments) to support and then oppose the thought to be examined.

4.1.3. Measures

1. Primary measures

Perceived Usefulness: This was a measure of how the participants perceived the usefulness of completing the thought-strengthening exercises. This was measured through a version of the usefulness subscale (six questions) from the Intrinsic Motivation Inventory [83]. The usefulness survey recorded participants' answers using a five-point Likert scale (1: "Strongly disagree"; 5: "Strongly agree") for each item. The questionnaire can be found in Appendix C.2.3.

Self-efficacy: The self-efficacy questionnaire was inspired by the work of [42], [118], [99] and [67]. All four studies followed Bandura [7] approach to produce their self-efficacy assessments based on situations, and we followed the approach to formulate a five-item questionnaire that was evaluated using a scale from 0 (Highly certain cannot do) to 10 (Highly certain can do), based on the suggestion by Bandura [7]. The self-efficacy questionnaire used in the experiment can be found in Appendix C.2.4.

2. Exploratory measures

Enjoyment was defined as the participants' enjoyment of overcoming the barrier during the experiment. The enjoyment questionnaire compared two conditions: the chatbot-support and the text-support groups. This was because it was related to their enjoyment of the barrier-overcoming method, which the third group (i.e., no support) did not receive. To measure this, we used only a post-experiment questionnaire (7 questions), derived from the Intrinsic Motivation Inventory [83], and asked participants to evaluate the items using a five-point Likert scale (1: "Strongly disagree"; 5: "Strongly agree") for each item. Additionally, an alternative measure for enjoyment was used, asking the participants if they, as Amazon Mechanical Turk workers, wanted to distribute 20 minutes of their time between two tasks: a picture-identification task and a thought-strengthening exercise. How much of the 20 minutes were they willing to spend on each task? This question was asked before and after participants completed the exercise in all three conditions.

Engagement was defined as the participants' engagement with the exercise. This was measured through active writing time, which was determined based on the time the participant spent interacting with the argument-writing exercise. If the participant did not interact with the exercise tab or did not do any activity (mouse movements, keyboard clicks, etc.) for 15 consecutive seconds, the timer was paused until he or she interacted again.

For the **believability differences in thoughts and barriers**, we asked participants to rate the believability of motivational barriers (to doing the exercise) and of the negative and positive thoughts. This was done before and after the experiment, and the rating was from 0 to 100.

4.1.4. Participants

Before starting the experiment, we decided to adjust our target group to be healthy people instead of PhD candidates specifically. To gain more concrete results within the time frame, a high number of participants, in addition to a quick way of gathering data, was needed. The first option was to conduct the test with PhD candidates who have shown depression symptoms. However, this was not feasible due to ethical constraints and uncertainty of willingness to take part in the evaluation. Therefore, we had two options to consider: either running the evaluation with healthy PhD students on campus, or conducting the experiment online with healthy people. We opted for the second option, as we could not guarantee more than 5 candidates, who are not familiar with the system, to be participants. Additionally, the data gathering could involve a lot of time to find suitable time slots for the students to perform the experiment (which was the reason for delaying the focus group session during the requirements-gathering phase). An additional reason for not including PhD students is that they might write arguments in the exercise by which they could be identified, which could be a crucial ethical constraint. Therefore, the second option was more feasible. The acquired analysis can be further used to examine the effect on motivation and to generalize the findings across another group (i.e., the PhD candidates).

A total of 174 participants (58 in each group) were recruited via the crowdsourcing platform Amazon Mechanical Turk. The number was determined by conducting a power analysis using G*Power. We based it on the difference between two independent-means t-tests to obtain a medium effect size of 0.5 with a power of 0.80 and a 0.05 error probability [17]. Table 4.1 shows the pre-measurements of perceived usefulness and self-efficacy for the three conditions. Overall, the groups' pre-measurements did not differ significantly (all P values > .05).

Table 4.1: A comparison of the pre-measurements for the three groups: no support, text support and chatbot support.

Model comparison	$\tilde{\chi}^2$ **(degrees of freedom)**	**P value**
No Support vs Chatbot support		
Pre Perceived Usefulness	0.01 (1)	0.92
Pre Self-efficacy	0.52 (1)	0.47
Text Support vs Chatbot Support		
Pre Perceived Usefulness	0.3 (1)	0.58
Pre Self-efficacy	2.79 (1)	0.13
No Support vs Text Support		
Pre Perceived Usefulness	0.22 (1)	0.64
Pre Self-efficacy	0.03 (1)	0.87

No demographics were collected from the participants; however, because fluency in English was important, the workers were specified to be residing in an English-speaking country (US, Canada, UK, Australia). Each worker was paid $3.50 for successfully completing the experiment, which took approximately 30 minutes.

Furthermore, the data of 29 participants who completed the exercise were excluded from the experiment and not included in the total 174 participants. The reasons for exclusion were performing the experiment more than once (n = 9), where only the first evaluation recording completed by the participants were included in the analysis; rating all pre-believability as 0, so it could not be lower in the post-believability measure (n = 5); writing nonsensical answers to the task questions (n = 13); and having the exact same answers as other participants to the open-ended questions (n = 2). These data do not include participants filtered out during the experiment due to answering the comprehension checks wrong, recording the thoughts' believability in less than 30 seconds, or willingly dropping out during the experiment. This is because the data of the filtered-out participants were not registered.

4.1.5. Procedure

This test followed a similar procedure to the experiment defined in [112]. Through Amazon Mechanical Turk, the participants accepted the job and then were redirected to Qualtrics to start the experiment. The test included four phases. The first phase was the introduction, in which the participants were asked to complete an informed consent form. Additionally, they were briefed about the nature of the experiment and informed about the exclusion policies for responses. After that, the participants were shown a video that explained what thought-strengthening exercises are and why they are used. Then, the video showed an example of how to complete an argument-writing exercise using an artificial scenario that was not reused during the experiment. This was followed by comprehension check questions on their understanding of the instructions for completing the exercise. In the second phase, they were asked to fill out a pre-questionnaire that covered three measures: perceived usefulness, self-efficacy, and the enjoyment question for distributing 20 minutes between tasks. Afterward, they were asked to rate the three motivational barriers based on applicability themselves. The participants were then given one of two new, randomly selected scenarios. They were asked to role-play in that situation and then asked to rate the believability of three negative and three positive thoughts depending on the situation. The highest rated (negative and positive) thoughts were used in the exercise to write arguments for and against. In the third phase, the participants were randomly assigned to one of the three conditions, wherein the chatbot and text groups received support corresponding to each participant's top motivational barrier, while the no-support group received nothing. In the motivation-addressing cases, the strategies for tackling the barrier were shown in a random order. In the chatbot case, the users had to interact with the agent and confirm if they understood or needed more convincing, which would result in the next strategy being shown. In the text case, the whole text was shown, and the participants were asked to read it. The proceed button was enabled after participants had spent one minute on the text page. Figure

4.3 shows a comparison of barrier tackling (if not convinced of the rationale) in the two conditions. When the participants proceeded to the exercise, they were asked to provide evidence for and against the most believable negative thought to examine it, then to provide evidence for and against the most believable positive thought. After that, they were asked again to rate the believability of the thoughts and the barriers, to answer the questionnaire again, and to respond to the picture-identification question again. An additional survey was given to the chatbot- and text-support groups to assess enjoyment of the barrier-addressing method. Figure 4.4 summarizes the procedure and the stages of the experiment.

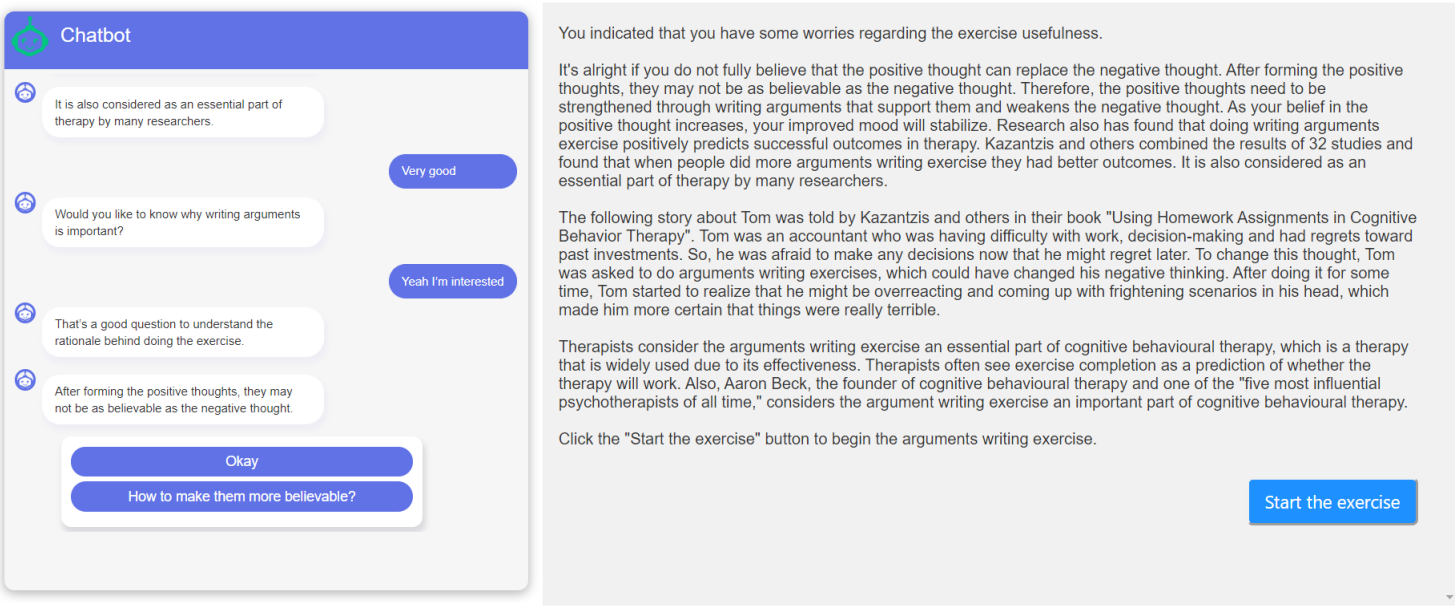


Figure 4.3: Motivational barrier tackling through the chatbot (left) and text (right)

4.1.6. Data preparation and statistical analysis

The analysis of the data was conducted using R 3.6.2. The R script markdown, along with the dataset, can be accessed online through 4TU data centre [93]. A reliability analysis was conducted for both measures: The self-efficacy scale had a Cronbach's alpha of 0.76, and perceived usefulness had a high alpha of 0.95, which is acceptable to use as an average according to Loewenthal and Lewis [80].

The experimental setup had a mixed design, so we used a multilevel model to test the hypotheses. Before fitting the data, we transformed the structure to fit a generalized model (ID, group, moment of measure, and perceived usefulness or self-efficacy scores). The gathered data were not linearly distributed; therefore, we used a generalized linear mixed-effect model [84]. Additionally, the scores were inverted so the residuals would fit gamma distribution, which showed no clear deviation from normality in P-P and Q-Q plots for both measures (perceived usefulness and self-efficacy). The models were computed using maximum likelihood. The models were computed using maximum likelihood with the participants as a random intercept effect and the pre- and post-measure scores as fixed effects. In other words, the models assumed that all participants would have the same pre- and post-change direction but with various starting points. The model was used to reveal any general differences in the measures' scores between the parameters in all three conditions (i.e., whether scores differed for the groups in general, for the moment of measurements, or for the interaction of both). To explore the found interaction effect between the groups, one group's data were systemically left out to check the other two groups and to obtain the comparison results. After that, we applied Spearman's rank correlation to check for a correlation between the delta between the post- and pre-scores of self-efficacy and perceived usefulness.

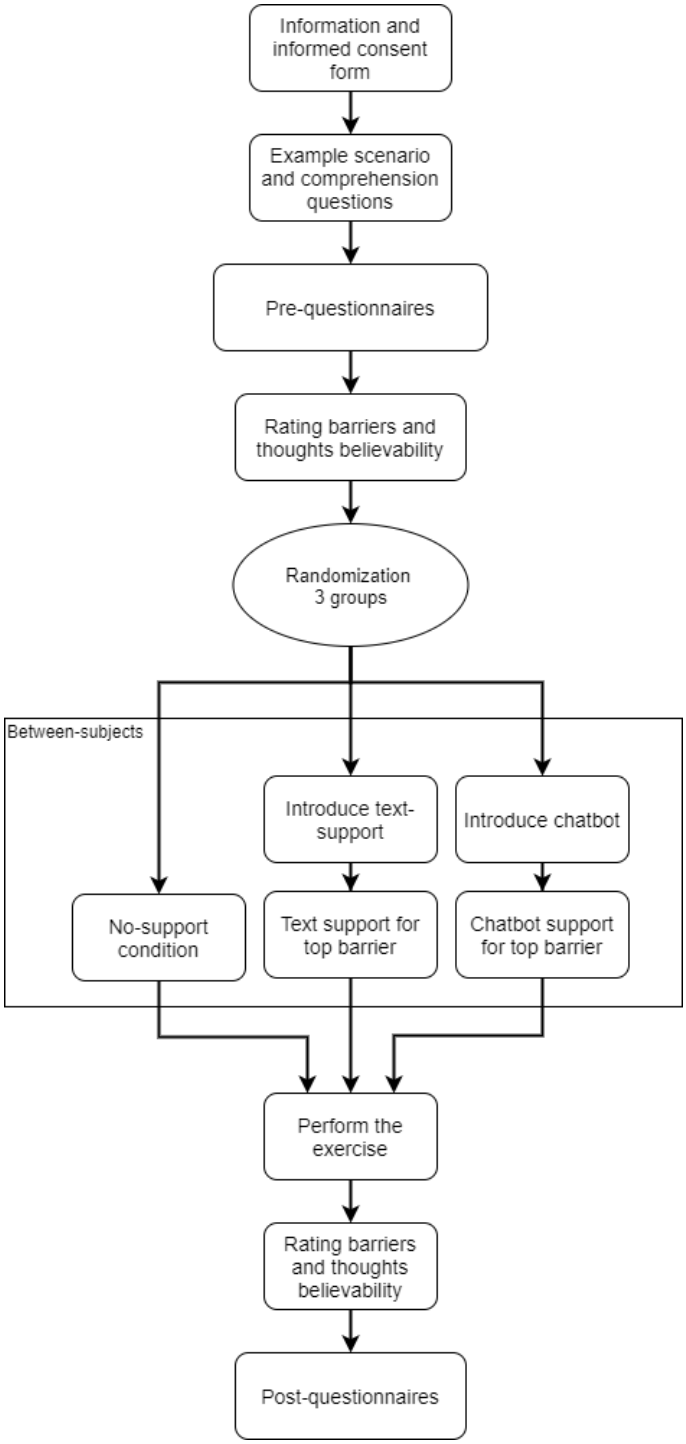


Figure 4.4: Outline of the experiment procedure.

4.2. Results

4.2.1. Perceived Usefulness

Table 4.2: Model comparisons for the effect of the support types on perceived usefulness.

Model comparison	Estimate	t value	P value
No Support vs Chatbot Support	0.055	3.216	0.001 **
Added chatbot support			
Text Support vs Chatbot Support	0.031	1.801	0.07
Added chatbot support			
No Support vs Text Support	0.023	1.346	0.17
Added text support			

When considering the support type effect and the moment of measure (i.e., pre and post) separately, the model did not show a significant difference of either, $\tilde{\chi}^2(2, N = 348) = 0.9$, $p > .05$ and $\tilde{\chi}^2(1, N = 348) = 1.99$, $p > .05$, respectively. However, when considering the interaction between both parameters, a significant association between the interaction and the measure scores could be noticed, $\tilde{\chi}^2(2, N = 348) = 10.38$, $p < .01$. Therefore, there was a difference in the perceived usefulness between the groups.

A comparison of the groups' effect on the perceived usefulness score can be seen in Table 4.2. The table reveals a significant difference between no support and chatbot support. However, there is no significant difference when considering the other data comparisons. The interaction effect can be seen between no support and chatbot support in Figure 4.5, which shows the measure of perceived usefulness for the three conditions and compares the change in the means of the pre-measures and the post-measures.

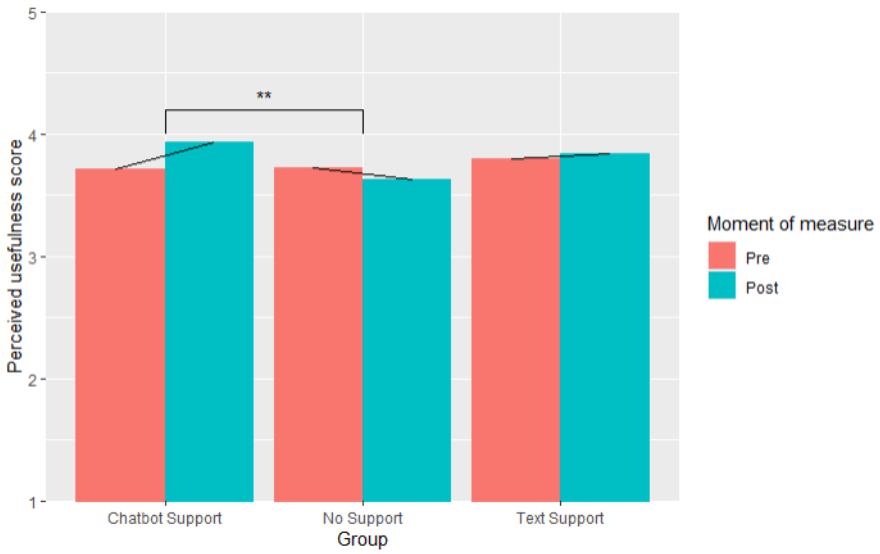


Figure 4.5: Bar chart comparing the mean of the pre and post scores for the perceived usefulness measure between the three conditions.

4.2.2. Self-efficacy

No significance is observed when considering the group effects on their own $\tilde{\chi}^2(2, N = 348) = 2.61$, $p > .05$. On the other hand, the pre- and post-session show a significant difference between their self-efficacy scores $\tilde{\chi}^2(1, N = 348) = 6.72$, $p < .01$. However, this is not the case regarding the interaction between the groups and the moment of measure, which indicates no significant difference $\tilde{\chi}^2(2, N = 348) = 2.75$, $p > .05$. Table 4.3 shows an overview of the means for the three conditions. Overall, an increase in self-efficacy appears for all three conditions.

Table 4.3: Mean and standard deviation results for the three conditions in the self-efficacy measure.

	Pre		Post	
Condition	**Mean**	**SD**	**Mean**	**SD**
No support	5.38	2.08	5.47	2.16
Text Support	5.22	1.77	5.41	2.03
Chatbot Support	5.75	2.01	6.11	2.05

For the correlation between the delta of the perceived usefulness and delta of the self-efficacy measures, we conducted a Spearman's test. The results indicate a positive correlation between the measures' Rs = 0.23, $p < 0.01$. This means that the more a participant's perceived usefulness increased, the more likely that self-efficacy also increased.

4.2.3. Explorative measures

In addition to the hypotheses, we looked at other explorative measures. Those measures, however, did not yield significant differences. Table 4.4 summarizes the results from the test, including the means and standard deviations of the conditions. The means of all measurements, except for the enjoyment questionnaire and the engagement time, were calculated from the delta of the measure (i.e., the difference between the post- and pre-measurements).

Table 4.4: The mean and standard deviation results for the explorative measurements. The table also shows model comparison of the measures, using a Kruskal-Wallis rank sum test for the conditions.

	No Support		Text Support		Chatbot Support			
Measurement	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	$\tilde{\chi}^2$ **(df)**	**P value**
Enjoyment:								
Questionnaire	N/A	N/A	2.99	0.55	3.11	0.57	2.1 (1)	0.14
Picture identification question	0.9	4.6	2.1	4.1	1.76	4.99	3.7 (2)	0.15
Engagement time	380	197	401	268	387	245	0.77 (2)	0.68
Believability of:								
Barriers	14.5	21.6	13.9	23.4	10.9	22.5	0.19 (2)	0.9
Thoughts	19.9	27.2	10.9	22.6	12.5	28.4	3.19 (2)	0.21

4.3. Discussion

The effect of the motivational content on perceived usefulness was confirmed in the results, especially in the chatbot condition. However, no clear effect was found for the presentation mode. These results partially confirm the first hypothesis (**H1**), as the participants who perceived the exercise as useful showed a significant increase in the chatbot support condition (pre and post), as well as when comparing the condition with the no-support condition. On the other hand, the results of comparing text support with no-support or chatbot support conditions did not show an observable difference. Thus, there is no evidence that the text-support can increase the perceived usefulness. One possible explanation can be that we enforced a minimum of one minute on the textual motivation page before proceeding to the exercise. This may have led some participants to be more exposed to the text and therefore reading parts of it more than once, thus increasing their factual retention [10]. This is not the case for the chatbot condition as the participants could proceed to the exercise directly after they finished interacting with the chatbot. In general, the participant perceived the chatbot condition's usefulness as higher than no support.

The analysis of the results for the self-efficacy measure were inconclusive and thus did not support the second hypothesis (**H2**). Therefore, no significant impact was found between the three conditions regarding participants' self-efficacy toward the exercise. A possible explanation for that can be explained via the

social cognitive theory [6]. The theory includes "Mastery Experiences," or first-hand experiences with successfully doing something as a primary source of self-efficacy. The case is the same here, where participants' self-efficacy increased in the three cases. This explanation can be further supported by the analysis, which revealed a significant effect of the pre- and post-sessions alone on self-efficacy, but not when considering the interaction with the groups. The cause may be attributed to the ceiling effect [33], where the humanly possible scale cannot show a noticeable growth regarding healthy participants. Also, people with depression symptoms tend to have less self-efficacy and esteem than healthy people do [45], which also aligns with CBT exercises [21]. Another possible reason is that the exercise may be easy and clear to participants because it is a fundamental exercise for strengthening thoughts in the aforementioned self-help book. This may lead the participants to complete the exercise correctly without hassle, thus increasing the initial self-efficacy and causing a smaller self-efficacy increase in all three conditions regardless of the motivational support type.

Notably, the analysis showed a positive correlation between the increase in participants' perceived usefulness and self-efficacy. In other words, the more a participant's perceived usefulness increased, the more likely their self-efficacy also increased (taking into consideration that the chatbot condition has the highest perceived usefulness). However, this may be due to rating biases, where people who give high scores tend to do so across scales, which has nothing to do with the conditions.

4.3.1. Limitations

The system was designed in the context of PhD candidates as the target group. We expected a relationship cycle starting with being more motivated to do the thought-strengthening exercise, which would lead the individual to be more engaged with the exercise and thus find and write more arguments and answers in response. In general, this can lead to strengthening the believability of the positive thought. However, we expected this cycle to be more feasible with PhD students than with Amazon Mechanical Turk workers, as the workers would most likely have a different primary incentive for completing the exercise (i.e., to be paid after completing the task). This may be a reason for not finding significant results for the exploratory measures, which are the components of the cycle.

Regarding the provided materials, the participants were given a defined scenario (that was deemed as widely relatable compared to other scenario options). The participants were also provided with the negative and positive thoughts for these scenarios. We could not expect all participants to relate to a scenario and to one negative and positive thought. An ideal solution to this limitation was to let users with depression symptoms record their thoughts on a real-life situation that made them feel negatively. Participants would then complete the exercise and record the measures. However, the ethical concerns of using an untested solution in this case was a crucial constraint.

One limitation of the no-support condition is that there is no content to substitute the motivation component found in the text and chatbot supports. In the no-support case, the participant started the exercise directly after rating the thoughts. This may have had an affect on the condition as it was not clear if the improvement with the motivation was due to the motivational content or was simply due to giving personal attention to the participants [73]. A solution could be to include a fourth condition that is similar to the no-support condition but includes giving attention as an element before the exercise. Although we cannot rule out this possibility, in real life, the exercises are mostly completed either with the motivational element (including the attention giving) or without the motivational element.

5

Discussion and Conclusion

In the thesis's final chapter, the research question answers are summarized. Next, the prototype implementation's limitations and evaluation are listed. This is followed by a discussion of this work's scientific and practical contributions. Finally, recommendations for the future are discussed.

5.1. Conclusion

This research primarily aimed to answer the following question:

How can a conversational intervention agent be used to motivate PhD candidates in changing negative thinking patterns?

This research question was divided into three sub questions:

*1- What are this conversational agent's **requirements**?*

At first, we investigated the primary issue that needed a solution: compared to the general public, PhD candidates are more vulnerable to depression-inducing situations, which explains the higher number of PhD students with depression symptoms. Therefore, cognitive behavioral therapy (CBT) is a widely used therapy to change negative thinking patterns into positive ones. One fundamental step of CBT is thought-strengthening exercises. However, therapists usually reported a lack of compliance from patients regarding the exercises. Therefore, a tool to motivate PhD students to perform thought-strengthening exercises is needed.

To meet the requirements, the candidates' values, human-factor knowledge, and related technology were examined and discussed, followed by focus groups sessions with PhD psychologists and PhD candidates. These activities were used to construct a list of requirements that can be seen as either process or value requirements. The process requirements follow steps that tackle motivational barriers to completing thought-strengthening exercises in real life. Those steps were: help students do the exercise, identify and address motivational barriers, provide psychoeducation to complete the exercise, and include means to perform the exercises. On the other hand, the values-based requirements were: understand and interact with PhD candidates, include transparency in privacy policies, and offer ease of access to the system.

*2- What **design** would meet the requirements?*

The system primarily aims to help PhD candidates overcome motivational barriers to performing thought-strengthening exercises. Therefore, the agent starts by identifying the student's motivational barrier to address it with appropriate motivational strategies. The system also includes the exercises themselves, so students can answer the homework questions. To facilitate chatbot interactions and the exercises, both windows (i.e., the chatbot and the exercises) appear beside each other and can be interacted with simultaneously . The conversational agent flow design was based on a real life sequence of addressing motivational barriers in a context of noncompliance with the thought-strengthening exercises. The chatbot also informs the candidates about the privacy policies and exercise-related information, which can be necessary to complete homeworks.

This textual conversational agent interacts with the PhD candidate through closed ended-questions and was implemented using Rasa NLP.

*3- Can such a conversational agent increase doctoral students' **motivation** toward thought-strengthening exercises?*

The prototype was evaluated with healthy individuals via Amazon Mechanical Turk, and covered two measures: participants' perceived usefulness and self-efficacy regarding completing the thought-strengthening exercises. The results showed a significant increase in the perceived usefulness of the chatbot condition before and after interacting with the agent and when comparing the condition with the no-support condition. Compared to text support, however, participants' perceived usefulness showed no clear variation. On the other hand, the testing revealed no noticeable differences between the three conditions when considering self-efficacy as a measure. However, the results showed that the participants' self-efficacy increased after completing the exercise in all conditions. Overall, the results imply that the motivational support could increase PhD candidates' motivation to complete thought-strengthening exercises because perceived usefulness is one motivational element in the Intrinsic Motivation Inventory [83].

5.2. Limitations

Some limitations were encountered during the course of this project. The first limitation is regarding lack of access to PhD students to reflect upon the final design of the system and act as participants in the experiment. Doctoral students' values and needs are different from those of the general public, and their feedback on the system might have proven beneficial if it had been acquired. The number of investigated motivational barriers is also a limitation that narrowed our scope. In Chapter 2, nine different barriers toward completing the exercises were identified. However, only four of these barriers were included in the design. Covering all nine barriers would help in addressing most issues and testing further; however, this would require additional time and resources to be examined and added to the system. Third is the focus on thought-strengthening exercises while designing the system, which usually are one step in the CBT pipeline. Restricting the design to one kind of exercise limited the approach from being more flexible. Additionally, because the chatbot was implemented via closed-ended interactions, the user had to follow predefined dialogue trees presented by the agent. This may restrict users to choosing from a predefined list of answers rather than writing what they have in mind, minimizing their sense of control [71].

As for the evaluation, the participant demographics (e.g., age and gender) were not recorded, which limited us in acquiring information on how the results could be generalized to other populations. Furthermore, the short-term setup of the experiment made it difficult to gather data regarding adherence and long-term benefits. As Gaynor et al. [56] found, the participants' overall compliance with the CBT homework decreased the further they advanced in the sessions, where the percentage of exercise completion dropped to less than half during the third and last set of sessions. Much relevant literature that used conversational agents in the context of CBT involved multi-week experiments rather than one session to establish the agents' effect (e.g., [51], [65]). However, this was not feasible in this study due to time constraints.

5.3. Contribution

On a scientific level, this research presented a novel approach of using a conversational agent to motivate PhD students in doing thought-strengthening exercises. As has been discussed, the topic of using technology with CBT homework has not been extensively explored. The research aimed at tackling the motivational barriers of PhD candidates toward completing thought-strengthening exercises. PhD students are usually more susceptible to depressive-inducing situations, which are not adequately addressed in eHealth systems. On a practical level, the system has the potential to be used in situations that may induce negative thoughts; however, this requires prior knowledge of how to apply CBT in general. Therefore, it could be suitable to use between therapy sessions or alongside self-help books to complete the thought-strengthening exercises. The evaluation of the system revealed a potential to increase PhD students' motivation toward thought-strengthening exercises.

5.4. Future work

A future possibility to widen the system's scope is to cover other mental health issues using similar types of exercises. For example, Greenberger and Padesky [59] advised people with anxiety symptoms to follow the same homework as those who have depression symptoms. On the other hand, the authors gave people with other disorders (e.g., panic attacks) a similar set of exercises in a different order. Therefore, choosing which exercise to start (i.e., determined by users or therapists) offered more flexibility and may be an approach to covering additional mental disorders using the same tool. Furthermore, additional CBT exercises can be explored and added to the system, bearing in mind that they may introduce motivational barriers other than the current barriers extracted from the thought-strengthening exercises. In addition, the system's target group can be broadened to include young adults, because PhD students and young adults share some core barriers [109].

The system can be extended further to enrich the provided support (e.g., adding a "hints" function to the exercise questions, where the chatbot gives suggestions and tips on how to answer the concerned question). This could also be accompanied with written examples and visual aids that can explain a situational example and how a character answered the questions in the exercise based on that situation. This method of explaining situations and characters was demonstrated by various systems (e.g., deprexis [86] and MoodGYM [24]); however, the systems involved web sessions rather than chatbots. Furthermore, adding the possibility to view the previously completed exercises can be beneficial for learning. Another idea stemming from the last point is to implement an additional interface for the psychologists so they can access their clients' exercises and check their progress. This would help in filling the gap that we did not include in the model of the CBT homework completion, which was the "reviewing" the homework step [68]. Another possibility to consider is to add praise for completing the exercise, as Fogg [52] argued that receiving compliments from a system can be comparable to receiving them from humans in terms of positive effects.

In Chapter 2, we identified several motivational strategies that cCBTs used (Table 2.3). During the PhD students' focus groups, some of these strategies' validities were discussed by the participants (see focus groups results, Section 2.4.2). One idea that might help increase users' adherence to the system is adding progress monitoring, where a new window can show how many exercises were completed and the extent of completion for each exercise, such as used in Help4Mood [22]. Another strategy for increasing adherence is to include a reward system, which can be implemented by adding a points system based on how many exercises a user completed. However, these functionalities need further testing for validity.

Other possible design approaches can be examined for the system. One approach is to identify the PhD candidates' motivational barriers through guided questions. The chatbot can navigate through a tree of questions to determine the most probable reason behind lack of motivation. Another approach is to change the core design by uniting the exercise and the chatbot, so the chatbot asks the exercise questions within the conversations, rather than having the exercise in a separate window.

As discussed in the limitations, conducting a multi-week evaluation may provide more accurate results, not just for adherence but also for other measures such as self-efficacy and enjoyment. This would result in exposing the participants to different types of exercises, as using one exercise may not be the best option to show the effect on self-efficacy as a whole.

5.5. Final remarks

In this thesis, a possible approach was presented for motivating PhD candidates with depression symptoms into doing thought-strengthening exercises. The evaluation results indicate that the support has the potential to raise the candidates' motivation to do the exercises. CBT homework compliance is a major issue for many, and this tool could provide a means to help the candidates overcome their motivational barriers to completing the exercises.

Bibliography

[1] Joy: Become your strongest self. URL `https://www.hellojoy.ai/support`.

[2] Arthur N Applebee and Judith A Langer. Instructional scaffolding: Reading and writing as natural language activities. *Language arts*, 60(2):168–175, 1983.

[3] Timothy B Baker, David H Gustafson, and Dhavan Shah. How can research keep up with ehealth? ten strategies for increasing the timeliness and usefulness of ehealth research. *Journal of medical Internet research*, 16(2):e36, 2014.

[4] David Bakker, Nikolaos Kazantzis, Debra Rickwood, and Nikki Rickard. Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR mental health*, 3 (1):e7, 2016.

[5] Albert Bandura. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84 (2):191, 1977.

[6] Albert Bandura. Social cognitive theory of personality. *Handbook of personality*, 2:154–196, 1999.

[7] Albert Bandura. Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents*, 5(1): 307–337, 2006.

[8] Albert Bandura. Self-efficacy. *The Corsini encyclopedia of psychology*, pages 1–3, 2010.

[9] Albert Bandura, WH Freeman, and Richard Lightsey. Self-efficacy: The exercise of control, 1999.

[10] Jerrold E Barnett and Richard W Seefeldt. Read something once, why read it again?: Repetitive reading and recall. *Journal of Reading Behavior*, 21(4):351–360, 1989.

[11] Judith S Beck and Aaron T Beck. *Cognitive therapy: Basics and beyond*. Number Sirsi) i9780898628470. Guilford press New York, 1995.

[12] Luciana Benotti, María Cecilia Martínez, and Fernando Schapachnik. Engaging high school students using chatbots. In *Proceedings of the 2014 conference on Innovation & technology in computer science education*, pages 63–68. ACM, 2014.

[13] Robbert Jan Beun, Siska Fitrianie, Fiemke Griffioen-Both, Sandor Spruit, Corine Horsch, Jaap Lancee, and Willem-Paul Brinkman. Talk and tools: the best of both worlds in mobile user interfaces for e-coaching. *Personal and ubiquitous computing*, 21(4):661–674, 2017.

[14] Timothy Bickmore, Amanda Gruber, and Rosalind Picard. Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient education and counseling*, 59(1): 21–30, 2005.

[15] Matthew D Blagys and Mark J Hilsenroth. Distinctive activities of cognitive–behavioral therapy: A review of the comparative psychotherapy process literature. *Clinical Psychology Review*, 22(5):671–706, 2002.

[16] Petter Bae Brandtzaeg and Asbjørn Følstad. Why people use chatbots. In *International Conference on Internet Science*, pages 377–392. Springer, 2017.

[17] Willem-Paul Brinkman. Design of a questionnaire instrument. *Handbook of mobile technology research methods*, pages 31–57, 2009.

[18] Willem-Paul Brinkman, VP Buil, R Cullen, R Gobits, and Floris L Van Nes. Design and evaluation of online multimedia maintenance manuals. *Behaviour & Information Technology*, 20(1):47–52, 2001.

[19] Lorraine Brown and Pamela Watson. Understanding the experiences of female doctoral students. *Journal of Further and Higher Education*, 34(3):385–404, 2010.

[20] Franziska Burger, Mark A Neerincx, and Willem-Paul Brinkman. Technological state of the art of electronic mental health interventions for major depressive disorder: Systematic literature review. *Journal of Medical Internet Research*, 22(1):e12599, 2020.

[21] David D Burns and Susan Nolen-Hoeksema. Coping styles, homework compliance, and the effectiveness of cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology*, 59(2):305, 1991.

[22] Christopher Burton, Maria Klara Wolters, Antoni Serrano Blanco, Aurora Szentagotai, Jenny Ure, Claudia Pagliari, and Brian McKinstry. Help4mood: avatar-based support for treating people with major depression in the community. *International journal of integrated care*, 12(Suppl1), 2012.

[23] Christopher Burton, Aurora Szentagotai Tatar, Brian McKinstry, Colin Matheson, Silviu Matu, Ramona Moldovan, Michele Macnab, Elaine Farrow, Daniel David, Claudia Pagliari, et al. Pilot randomised controlled trial of help4mood, an embodied virtual agent-based system to support treatment of depression. *Journal of telemedicine and telecare*, 22(6):348–355, 2016.

[24] Alison L Calear, Helen Christensen, Andrew Mackinnon, and Kathleen M Griffiths. Adherence to the moodgym program: outcomes and predictors for an adolescent school-based population. *Journal of affective disorders*, 147(1-3):338–344, 2013.

[25] Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. Towards a chatbot for digital counselling. In *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*, page 24. BCS Learning & Development Ltd., 2017.

[26] Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. Best practices for designing chatbots in mental healthcare–a case study on ihelpr. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*, pages 1–5, 2018.

[27] Justine Cassell, Joseph Sullivan, Elizabeth Churchill, and Scott Prevost. *Embodied conversational agents*. MIT press, 2000.

[28] Dianne L Chambless and Thomas H Ollendick. Empirically supported psychological interventions: Controversies and evidence. *Annual review of psychology*, 52(1):685–716, 2001.

[29] Victoria L Champion, Celette Sugg Skinner, et al. The health belief model. *Health behavior and health education: Theory, research, and practice*, 4:45–65, 2008.

[30] Robert B Cialdini and Robert B Cialdini. Influence: The psychology of persuasion. 1993.

[31] Leon Ciechanowski, Aleksandra Przegalinska, Mikolaj Magnuski, and Peter Gloor. In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92:539–548, 2019.

[32] David A Cook and Anthony R Artino Jr. Motivation to learn: an overview of contemporary theories. *Medical education*, 50(10):997–1014, 2016.

[33] Duncan Cramer and Dennis Laurence Howitt. *The Sage dictionary of statistics: a practical resource for students in the social sciences*. Sage, 2004.

[34] Remmert Daas, Gert-Jan Munneke, Heather Bray, Monalisa Goswami, and Jort ten Berg. *UvAPro PhD Survey 2015: Are UvA PhD candidates at risk of depression?* 2016.

[35] Fred D Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340, 1989.

[36] Victor Day, Patrick J McGrath, and Magdalena Wojtowicz. Internet-based guided self-help for university students with anxiety, depression and stress: a randomized controlled clinical trial. *Behaviour research and therapy*, 51(7):344–351, 2013.

[37] Edward L Deci. Effects of externally mediated rewards on intrinsic motivation. *Journal of personality and Social Psychology*, 18(1):105, 1971.

[38] Sebastian Deterding. Gamification: designing for motivation. *interactions*, 19(4):14–17, 2012.

[39] Jerusha B Detweiler and Mark A Whisman. The role of homework assignments in cognitive therapy for depression: Potential methods for enhancing adherence. *Clinical Psychology: Science and Practice*, 6 (3):267–282, 1999.

[40] Todd Dewett, Shung Jae Shin, Soo Min Toh, and Matthew Semadeni. Doctoral student research as a creative endeavour. *College Quarterly*, 8(1):n1, 2005.

[41] Marianne Di Pierro. Excellence in doctoral education: Defining best practices. *College Student Journal*, 41(2), 2007.

[42] Ding Ding, Willem-Paul Brinkman, and Mark A Neerincx. Simulated thoughts in virtual reality for negotiation training enhance self-efficacy and knowledge. *International Journal of Human-Computer Studies*, page 102400, 2020.

[43] Rachael Dyson and Kimberly Renk. Freshmen adaptation to university life: Depressive symptoms, stress, and coping. *Journal of clinical psychology*, 62(10):1231–1244, 2006.

[44] Ruth E Edelman and Dianne L Chambless. Compliance during sessions and homework in exposure-based treatment of agoraphobia. *Behaviour research and therapy*, 31(8):767–773, 1993.

[45] Marion F Ehrenberg, David N Cox, and Raymond F Koopman. The relationship between self-efficacy and depression in adolescents. *Adolescence*, 26(102):361, 1991.

[46] A. Engberg. Mental health chatbots—the future of therapy?, 2017.

[47] Ahmed Fadhil and Silvia Gabrielli. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. In *Proceedings of the 11th EAI international conference on pervasive computing Technologies for Healthcare*, pages 261–265. ACM, 2017.

[48] Ahmed Fadhil and Adolfo Villafiorita. An adaptive learning with gamification & conversational uis: The rise of cibopolibot. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization*, pages 408–412. ACM, 2017.

[49] Sina Falaki. Chat bot etiquette, Oct 2016. URL `https://medium.com/chatbotglobalnews/chat-bot-etiquette-5db39f401f30`.

[50] Andy Field and Graham Hole. *How to design and report experiments*. Sage, 2002.

[51] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 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):e19, 2017.

[52] Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002 (December):2, 2002.

[53] Brian J Fogg. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*, pages 1–7, 2009.

[54] National Collaborating Centre for Mental Health (UK et al. Depression: the treatment and management of depression in adults (updated edition). British Psychological Society, 2010.

[55] Hannah Gaffney, Warren Mansell, and Sara Tai. Conversational agents in the treatment of mental health problems: Mixed-method systematic review. *JMIR Mental Health*, 6(10):e14166, 2019.

[56] Scott T Gaynor, P Scott Lawrence, and Rosemery O Nelson-Gray. Measuring homework compliance in cognitive-behavioral therapy for adolescent depression: Review, preliminary findings, and implications for theory and practice. *Behavior modification*, 30(5):647–672, 2006.

[57] Rich Gilman, Jeffrey S Ashby, Dina Sverko, Dan Florell, and Kristen Varjas. The relationship between perfectionism and multidimensional life satisfaction among croatian and american youth. *Personality and individual differences*, 39(1):155–166, 2005.

[58] Ian H Gotlib and Constance L Hammen. *Handbook of depression*. Guilford Press, 2008.

[59] Dennis Greenberger and Christine A Padesky. *Mind over mood: Change how you feel by changing the way you think*. Guilford Publications, 2015.

[60] Rosanna E Guadagno, Jim Blascovich, Jeremy N Bailenson, and Cade McCall. Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology*, 10(1):1–22, 2007.

[61] Sylvia Helbig and Lydia Fehm. Problems with homework in cbt: Rare exception or rather frequent? *Behavioural and Cognitive Psychotherapy*, 32(3):291–301, 2004.

[62] Jennifer Hill, W Randolph Ford, and Ingrid G Farreras. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49:245–250, 2015.

[63] Simon Hoermann, Kathryn L McCabe, David N Milne, and Rafael A Calvo. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. *Journal of medical Internet research*, 19(8):e267, 2017.

[64] Jennifer L Hudson and Philip C Kendall. Showing you can do it: Homework in therapy for children and adolescents with anxiety disorders. *Journal of Clinical Psychology*, 58(5):525–534, 2002.

[65] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, 6(11):e12106, 2018.

[66] Eva Kaltenthaler, Glenys Parry, Catherine Beverley, and Michael Ferriter. Computerised cognitive-behavioural therapy for depression: systematic review. *The British Journal of Psychiatry*, 193(3):181–184, 2008.

[67] Ni Kang. Public speaking in virtual reality: Audience design and speaker experiences. 2016.

[68] Nikolaos Kazantzis, Frank P Deane, Kevin R Ronan, and Luciano L'Abate. *Using homework assignments in cognitive behavior therapy*. Routledge, 2005.

[69] Robert Kohn, Shekhar Saxena, Itzhak Levav, and Benedetto Saraceno. The treatment gap in mental health care. *Bulletin of the World health Organization*, 82:858–866, 2004.

[70] Lean L Kramer, Silke ter Stal, Bob C Mulder, Emely de Vet, and Lex van Velsen. Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review. *Journal of Medical Internet Research*, 22(2):e14058, 2020.

[71] Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini, Arianna Manzini, Ilina Singh, and NeurOx Young People's Advisory Group. 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*, 11:1178222619829083, 2019.

[72] Glenn W Lambie and Nicole Vaccaro. Doctoral counselor education students' levels of research self-efficacy, perceptions of the research training environment, and interest in research. *Counselor Education and Supervision*, 50(4):243–258, 2011.

[73] Peter J Lang. The emotion probe: studies of motivation and attention. *American psychologist*, 50(5): 372, 1995.

[74] Robert L Leahy, Stephen J Holland, and Lata K McGinn. *Treatment plans and interventions for depression and anxiety disorders*. Guilford press, 2011.

[75] Richard T LeBeau, Carolyn D Davies, Najwa C Culver, and Michelle G Craske. Homework compliance counts in cognitive-behavioral therapy. *Cognitive Behaviour Therapy*, 42(3):171–179, 2013.

[76] Eun Ju Lee, Clifford Nass, and Scott Brave. Can computer-generated speech have gender? an experimental test of gender stereotype. In *CHI'00 extended abstracts on Human factors in computing systems*, pages 289–290, 2000.

[77] James Lester, Karl Branting, and Bradford Mott. Conversational agents. *The Practical Handbook of Internet Computing*, pages 220–240, 2004.

[78] B Liu and SS Sundar. Should machines express sympathy and empathy? experiments with a health advice chatbot. *Cyberpsychology, Behavior and Social Networking*, 21(10):625–636, 2018.

[79] Edwin A Locke and Gary P Latham. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, 57(9):705, 2002.

[80] Kate Miriam Loewenthal and Christopher Alan Lewis. *An introduction to psychological tests and scales*. Psychology press, 2018.

[81] Amanda Longfield, Joanne Romas, and Jennifer D Irwin. The self-worth, physical and social activities of graduate students: A qualitative study. *College Student Journal*, 40(2):282–293, 2006.

[82] Kien Hoa Ly, Ann-Marie Ly, and Gerhard Andersson. A fully automated conversational agent for promoting mental well-being: a pilot rct using mixed methods. *Internet interventions*, 10:39–46, 2017.

[83] Edward McAuley, Terry Duncan, and Vance V Tammen. Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, 60(1):48–58, 1989.

[84] Charles E McCulloch and John M Neuhaus. Generalized linear mixed models. *Encyclopedia of biostatistics*, 4, 2005.

[85] Sally N Merry, Karolina Stasiak, Matthew Shepherd, Chris Frampton, Theresa Fleming, and Mathijs FG Lucassen. The effectiveness of sparx, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. *Bmj*, 344:e2598, 2012.

[86] Björn Meyer, Thomas Berger, Franz Caspar, Christopher Beevers, Gerhard Andersson, and Mario Weiss. Effectiveness of a novel integrative online treatment for depression (deprexis): randomized controlled trial. *Journal of medical Internet research*, 11(2):e15, 2009.

[87] Lisa N Michaud and Kathleen F McCoy. Capturing the evolution of grammatical knowledge in a call system for deaf learners of english. *International Journal of Artificial Intelligence in Education*, 16(1): 65–97, 2006.

[88] Joao Luis Zeni Montenegro, Cristiano André da Costa, and Rodrigo da Rosa Righi. Survey of conversational agents in health. *Expert Systems with Applications*, 2019.

[89] Maurice Mulvenna, Jennifer Boger, and Raymond Bond. Ethical by design: A manifesto. In *Proceedings of the European Conference on Cognitive Ergonomics 2017*, pages 51–54, 2017.

[90] Mark A Neerincx and Jasper Lindenberg. *Situated cognitive engineering for complex task environments*. Ashgate Publishing Limited Aldershot, 2008.

[91] DA Norman. Some observations on mental models. In *Human-computer interaction: a multidisciplinary approach*, pages 241–244. 1987.

[92] Anthony J Onwuegbuzie. Academic procrastinators and perfectionistic tendencies among graduate students. *Journal of Social Behavior and Personality*, 15(5; SPI):103–110, 2000.

[93] Mohammed Al Owayyed. Dataset of "motivating phd candidates with depression symptoms to complete thoughts-strengthening exercises via a conversational agent", 2020. URL `https://data.4tu.nl/repository/uuid:6ec331bc-8826-427f-9a12-6aff9a5849c8`.

[94] Mohammed Al Owayyed and Franziska Burger. Computer-based intervention for supporting individuals in changing negative thinking patterns. URL `https://osf.io/v6tkq`.

[95] Galen Panger, Janell Tryon, and Andrew Smith. *Graduate Student Happiness Well-Being Report | 2014*.

[96] Christian Paping, Willem-Paul Brinkman, and Charles van der Mast. An explorative study into tele-delivered multi-patient virtual reality exposure therapy system. *Coping with posttraumatic stress disorder in returning troops: Wounds of War II*, pages 203–219, 2010.

[97] Jacqueline B Persons, Joan Davidson, and Michael A Tompkins. *Essential components of cognitive-behavior therapy for depression.* American Psychological Association, 2001.

[98] E Lisa Price, Peter J McLeod, Stephen S Gleich, and Denise Hand. One-year prevalence rates of major depressive disorder in first-year university students. *Canadian Journal of Counselling and Psychotherapy/Revue canadienne de counseling et de psychothérapie*, 40(2), 2006.

[99] Chao Qu, Yun Ling, Ingrid Heynderickx, and Willem-Paul Brinkman. Virtual bystanders in a language lesson: examining the effect of social evaluation, vicarious experience, cognitive consistency and praising on students' beliefs, self-efficacy and anxiety in a virtual reality environment. *PloS one*, 10(4), 2015.

[100] Nicole M Radziwill and Morgan C Benton. Evaluating quality of chatbots and intelligent conversational agents. *arXiv preprint arXiv:1704.04579*, 2017.

[101] Jeff Riggenbach. *The CBT toolbox: A workbook for clients and clinicians.* PESI Publishing & Media, 2012.

[102] Juan Carlos Roca and Marylène Gagné. Understanding e-learning continuance intention in the workplace: A self-determination theory perspective. *Computers in human behavior*, 24(4):1585–1604, 2008.

[103] Joaquin Rodriguez, Gabriele Piccoli, and Marcin Bartosiak. Nudging the classroom: Designing a socio-technical artifact to reduce academic procrastination. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

[104] C Douglas Saddler and Laurie A Sacks. Multidimensional perfectionism and academic procrastination: Relationships with depression in university students. *Psychological Reports*, 1993.

[105] Marty Sapp. Three treatments for reducing the worry and emotionality components of test anxiety with undergraduate and graduate college students: Cognitive-behavioral hypnosis, relaxation therapy, and supportive counseling. *Journal of College Student Development*, 37(1):79–87, 1996.

[106] Stephen M Schueller, Kathryn Noth Tomasino, and David C Mohr. Integrating human support into behavioral intervention technologies: the efficiency model of support. *Clinical Psychology: Science and Practice*, 24(1):27–45, 2017.

[107] Abigail Sellen, Bill Buxton, and John Arnott. Using spatial cues to improve videoconferencing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 651–652, 1992.

[108] Øystein Sørebø, Hallgeir Halvari, Vebjørn Flaata Gulli, and Roar Kristiansen. The role of self-determination theory in explaining teachers' motivation to continue to use e-learning technology. *Computers & Education*, 53(4):1177–1187, 2009.

[109] Anna Sverdlik, Nathan C Hall, Lynn McAlpine, and Kyle Hubbard. The phd experience: A review of the factors influencing doctoral students' completion, achievement, and well-being. *International Journal of Doctoral Studies*, 13:361–388, 2018.

[110] Phillip J Tichenor, George A Donohue, and Clarice N Olien. Mass media flow and differential growth in knowledge. *Public opinion quarterly*, 34(2):159–170, 1970.

[111] Myrthe Tielman, Mark Neerincx, and Willem-Paul Brinkman. Generating situation-based motivational feedback in a ptsd e-health system. In *International Conference on Intelligent Virtual Agents*, pages 437–440. Springer, 2017.

[112] Myrthe L Tielman, Mark A Neerincx, and Willem-Paul Brinkman. Design and evaluation of personalized motivational messages by a virtual agent that assists in post-traumatic stress disorder therapy. *Journal of medical Internet research*, 21(3):e9240, 2019.

[113] Alexander JAM Van Deursen and Jan AGM Van Dijk. The digital divide shifts to differences in usage. *New media & society*, 16(3):507–526, 2014.

[114] Jan AGM Van Dijk. Digital divide: Impact of access. *The International Encyclopedia of Media Effects*, pages 1–11, 2017.

[115] David L Vogel, Stephen R Wester, and Lisa M Larson. Avoidance of counseling: Psychological factors that inhibit seeking help. *Journal of Counseling & Development*, 85(4):410–422, 2007.

[116] Paul Voigt and Axel Von dem Bussche. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 2017.

[117] MJ Wells, JJ Owen, LW McCray, LB Bishop, TD Eells, GK Brown, D Richards, ME Thase, and JH Wright. Computer-assisted cognitive-behavior therapy for depression in primary care: Systematic review and meta-analysis., 2018.

[118] Dongping Zheng, Michael F Young, Robert A Brewer, and Manuela Wagner. Attitude and self-efficacy change: English language learning in virtual worlds. *CALICO journal*, 27(1):205–231, 2009.

A

Foundation Appendices

A.1. Developed personas

Table A.1: A description of a persona who is Sarah.

Demographics	
Name	Sarah
Gender	Female
Age	31
Department	CEG
Years since the program started	2
Relationship status	Married
Number of children	2 (twins)
Children's ages	1 year old
Motivation	
Hobbies	Reading
University	Sarah is interested in her current research, although she thinks that she was better in her last job as a civil engineer As she would have more time to spend with her family.
Personal challenges	Sarah feels that her personal life is interfering with her academic life. Sometimes, she stays up all night to complete her assigned task, while her husband, Bob, put the children to bed.
Social Environment	
Background	Sarah is from a middle-class family. She is an international PhD candidate doing research in the Civil Engineering Department. She lives in an apartment with her husband, who is also a PhD candidate The university is a 15-minute bike ride from where she lives.
Supervisor & Colleagues	Sarah has casual relationships with her colleagues and supervisor, although she does not usually meet them outside working hours. She is a self-proclaimed introvert; therefore she is not an outgoing person.

Table A.2: A description of a persona who is Aaliyah.

Demographics	
Name	Aaliyah
Gender	Female
Age	29
Department	TPM
Years since the program started	1
Relationship status	Single
Motivation	
Hobbies	Basketball
University	Aaliyah is interested in her current research, although she thinks that her supervisor does not appreciate her hard work.
Personal challenges	Aaliyah feels that she is lost in her research and will not be able to finish. She tries to discuss this with her supervisor but she feels he does not invest the required time and effort in guiding her.
Social Environment	
Background	Aaliyah is from a high-class family. She is a PhD candidate doing research in Technology, Policy and Management. She lives in an apartment with her family, who have been living in the Netherlands for 20 years.
Supervisor & Colleagues	Aaliyah does not like her supervisor, and thinks about changing her research because of him. She likes to go out with friends and colleagues.

Table A.3: A description of a persona who is Alberto.

Demographics	
Name	Alberto
Gender	Male
Age	32
Department	EEMCS
Years since the program started	2
Relationship status	Married
Number of children	0
Motivation	
Hobbies	Playing video games
University	Alberto is interested in his current research, although he thinks that he is always not doing enough work.
Personal challenges	Alberto always thinks that he is not doing enough work. He based this thought on comparing his work to that of his colleagues and supervisor. He keeps trying to perfect the smallest aspects of his work.
Social Environment	
Background	Alberto is from a middle-class family. He is an international PhD candidate doing research in the Electrical Engineering Department. He lives in a studio with his wife.
Supervisor & Colleagues	Alberto has casual relationships with his colleagues and supervisor. Before starting his PhD program, Alberto was an outgoing person. However, when he spends time with his friends, he feels guilty for wasting his time.

A.2. Mock-ups

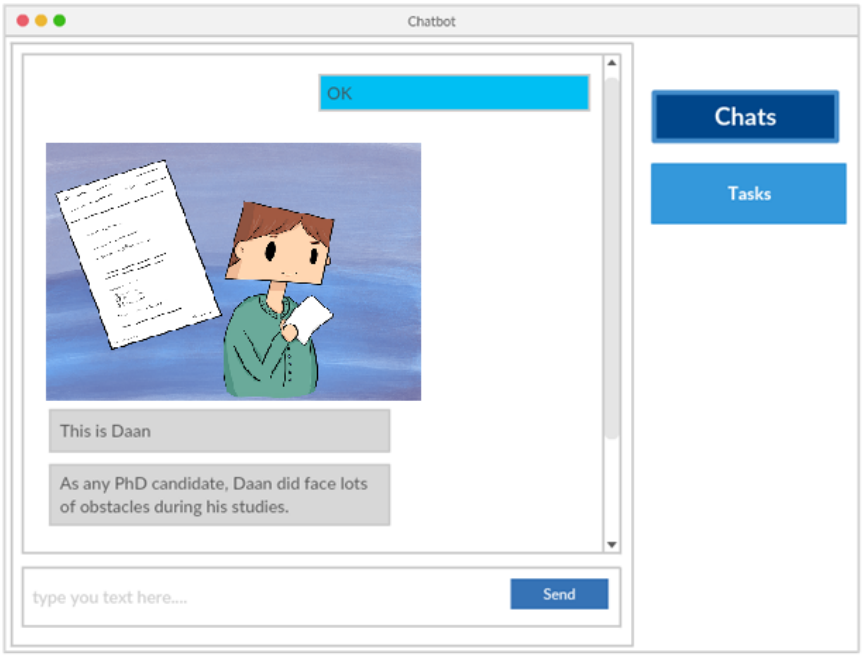


Figure A.1: A mock-up of how a chatbot can present a social model for addressing barriers.

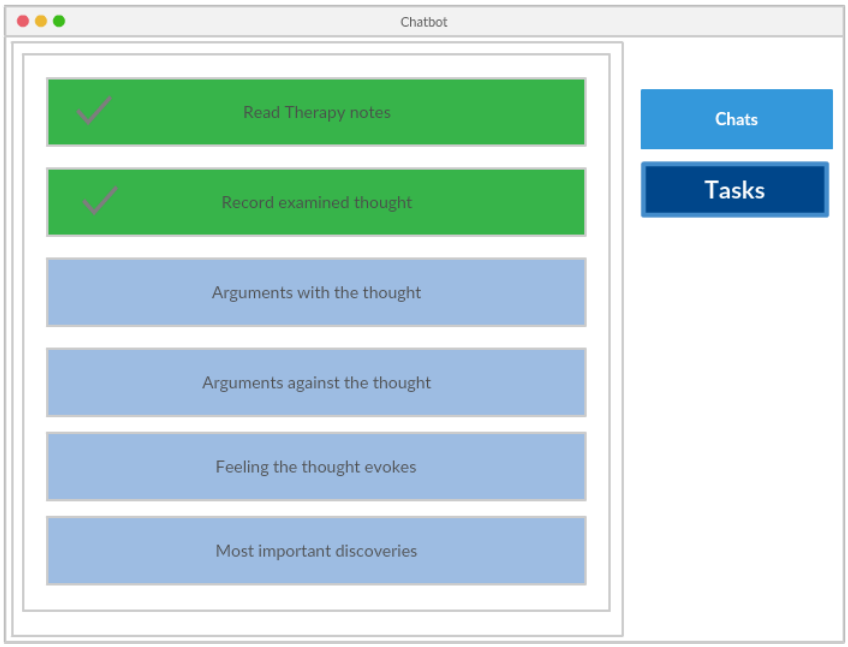


Figure A.2: A mock-up of how the exercise can be represented in the system.

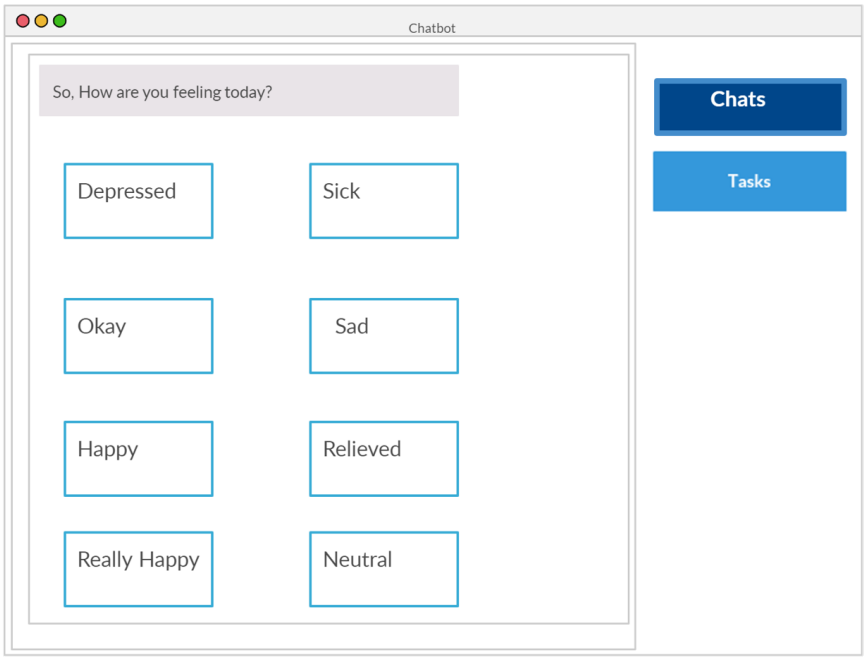


Figure A.3: A mock-up of how the closed-ended questions may work in the chatbot.

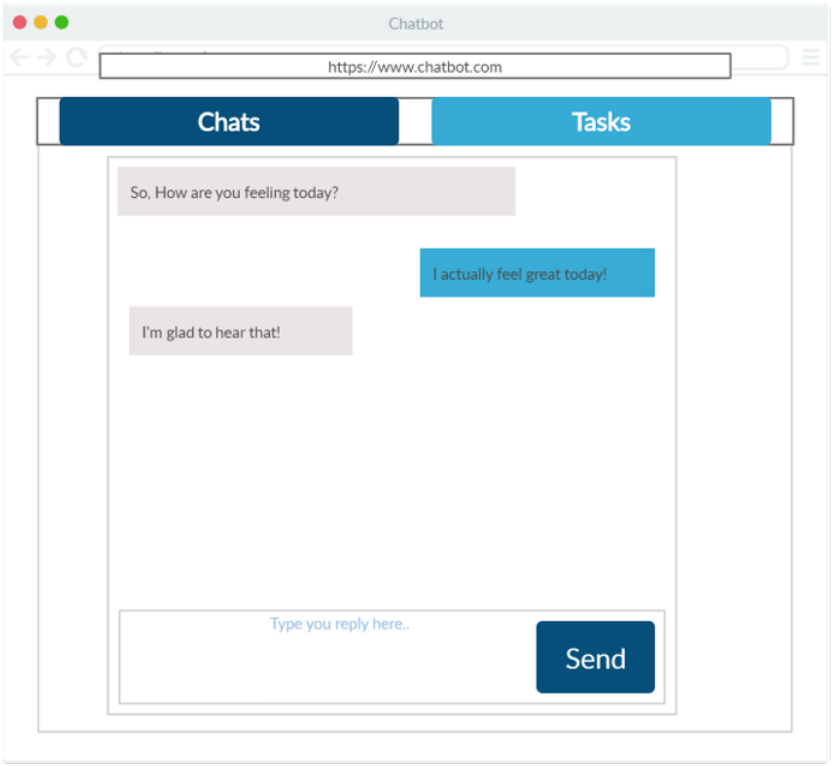


Figure A.4: A mock-up of how the system may look on a webpage.

A.3. Scenarios and suggestions

Table A.4: A scenario and the arguments derived from focus groups. [Psychologists]

Scenario 2: Not convinced	
Sarah is not convinced these exercises are useful. She finds it difficult to work on something with which she does not agree.	
Strategy A: Anecdotal	
Description	**Arguments supporting A:**
- The chatbot presents Sarah with a social model (i.e., a PhD student with similar motivational issues) in the form of a story. - The story shows the positive outcome of doing the exercises.	- In general, research suggests that concrete personal examples are more efficient than statistics. - This strategy is more accessible and less abstract. - This strategy is more personalized.
	Arguments opposing A:
	- The strategy might help only in some cases (e.g., those with similar demographics). - Many social models are necessary, which then must be matched to the client's issue. - The story should really fit the client. - The strategy relies a lot on the client's capacity for imagination. - The strategy would be better if the social model were a real person.
Strategy B: Studies and statistics	
Description	**Arguments supporting B:**
- The chatbot attempts to convince Sarah of the importance of completing the homework using studies and statistics. - The chatbot can link to additional relevant information (e.g., a YouTube video).	- For PhD candidates, studies and statistics are generally convincing. - This strategy is general and can be integrated with additional information.
	Arguments opposing B:
	- This strategy depends on the content of the video.
Suggestion	
Strategy B (statistics) would work better for PhD students. Finding a model that fits all students can prove difficult. In addition, defining numerous social models to choose from can lead to the need to collect more personal data. However, the social model can also be included in the linked video, in which a real social model can describe their story.	

Table A.5: A scenario and the arguments derived from focus groups. [Psychologists]

<table>
<tr><th colspan="2">Scenario 3: Low self-efficacy</th></tr>
<tr><td colspan="2">Aaliyah lacks the confidence to do the exercises. She thinks she is not capable of doing them and feels badly about herself.</td></tr>
<tr><th colspan="2">Strategy A: Verbal encouragement</th></tr>
<tr><th>Description</th><th>Arguments supporting A:</th></tr>
<tr><td rowspan="2">- The chatbot shows Aaliyah an overview of the completed tasks (past achievements).
- The chatbot uses exercises to raise self-efficacy, such as asking Aaliyah to list 3 things that went well last week.
- The tasks' structure does not change in this scenario.</td><td>- This strategy more directly and transparently addresses the low self-efficacy issue.
- Focusing on positive affirmation helps the client remain aware of the tasks they did well.</td></tr>
<tr><th>Arguments opposing A:</th></tr>
<tr><td></td><td>- This strategy could be less effective with more severe cases, as boosting mood is obvious and might increase resistance.
- The semi-positive events could be downsized, which would lead to increase feelings of failure.
- Negative cognitive bias could make it more difficult to remember positive events.</td></tr>
<tr><th colspan="2">Strategy B: Assist with tasks</th></tr>
<tr><th>Description</th><th>Arguments supporting B:</th></tr>
<tr><td rowspan="2">- The chatbot tries to assist Aaliyah in completing the tasks.
- Instead of showing one large task, the chatbot divides it into smaller ones to make it easier to complete.
- Aaliyah has to complete the small tasks, which will be combined into the main one.</td><td>- This strategy divides the tasks into manageable, concrete steps.
- Breaking up the tasks increases the client's chances of starting them.
- Completing small tasks gives the client a sense of success, so this strategy helps to grow confidence.</td></tr>
<tr><th>Arguments opposing B:</th></tr>
<tr><td></td><td>- This strategy has a short-term positive effect. However, it can enhance client's sense of incapability in the long run, thus lowering the self-efficacy more.</td></tr>
<tr><th colspan="2">Suggestion</th></tr>
<tr><td colspan="2">Although verbal encouragement could be less effective with severe cases, it provides a safer solution in the long run. However, the "list 3 things that went well" question should be changed to a more personal question (e.g., "List 3 things you did well last week").</td></tr>
</table>

Table A.6: A scenario and the arguments derived from focus groups. [Psychologists]

Scenario 4: Perfectionism	
Alberto strives to do the homework perfectly. He feels he must either complete it all correctly, or do nothing. This mentality hinders him from starting the assignments.	
Strategy A: Simple reminder	
Description	**Arguments supporting A:**
- The chatbot explains that the answers do not need to be perfect. - When Alberto accesses a task, a reminder about not answering everything perfectly is displayed.	- This strategy encourages the client to start the task without striving to be perfect.
	Arguments opposing A:
	- This strategy does not address perfectionism directly. Instead, it tries to convince the client to forget about it. - This strategy is milder and therefore less powerful. - People already know that not everything should be perfect. - Directing the client's attention to perfectionism could have opposite effects. - Some research evidence indicates that people notice the essential words (i.e., perfectionism) and overlook the additional "not."
Strategy B: Behavioral experiment	
Description	**Arguments supporting B:**
- The chatbot asks Alberto to deliberately make mistakes (e.g., grammatical errors or not doing the task thoroughly). - A time limit (countdown) is added for each task. - Alberto has to complete the task before the time ends, or the task will be submitted.	- This strategy includes a behavioral experiment to investigate the effects of doing something imperfectly. - The PhD students will experience the following: if they make a mistake: (a) nothing happens, (b) they will be given a chance to correct it, or (c) they will receive a comment about it. - This strategy includes nice self-competitive element, which positively correlates with successful performance.
	Arguments opposing B:
	- Submitting tasks automatically perhaps goes too far.
Suggestion	
Strategy B is clearly preferable. However, the aim should be more explicit at the beginning. Therefore, Strategy A will be used first, and if the problem persists, B will be used.	

Table A.7: A scenario and the arguments derived from focus groups. [PhD Candidates]

Scenario 5: Interactions	
How can Bill and the chatbot interact during the conversation?	
Strategy A: Multiple choices	
Description	**Arguments supporting A:**
- The chatbot presents predefined answers for its questions and interactions with Bill. - Bill must choose the most applicable answer.	- This strategy is faster. - This strategy is also more intuitive.
	Arguments opposing A:
	- No answers may be relevant.
Strategy B: Open answers	
Description	**Arguments supporting B:**
- The chatbot asks Bill to write their answer in text. - The chatbot analyzes the answer and replies based on its understanding.	- Open answers lead to better self-expression.
	Arguments opposing B:
	- The chatbot may not be good enough to detect the client's true intentions.
Suggestion	
Both strategies should be combined such that most communication is done through closed-ended questions. However, there can be an option for the PhD candidate to elaborate, after which the chatbot can confirm that it understood the PhD candidate's intentions.	

Table A.8: A scenario and the arguments derived from focus groups. [PhD Candidates]

Scenario 6: Platform	
How the system can be accessed by Bill?	
Strategy A: Standalone software	
Description	**Arguments supporting A:**
The software is installed on the client's device.	- This option is better for privacy.
	Arguments opposing A:
	- This option can be less accessible.
Strategy B: Webpage	
Description	**Arguments supporting B:**
The software can be accessed through a web browser.	- This option is more accessible for people who would like to try the system. - This option can be accessed on mobile devices.
	Arguments opposing B:
	- Conversation data are stored in a database.
Suggestion	
Hosting the chatbot on a secure website will make it more accessible to the students. It was also their preference, taking in mind that privacy policies should be transparent (e.g., how the data is handled, where it is stored, etc.)	

A.4. Minor Requirements

Table A.9: List of the systems' minor requirements.

9. Facilitate a sense of accountability.
10. Be able to convey empathy.
11. Encourage PhD candidates to perform tasks by giving them rewards.
12. Be visually appealing.
13. Have fluent conversations with the PhD candidate.
14. An option that lets PhD students elaborate on their answer should be present.
15. Monitor the progress of the user.
16. Use images and "emojis" in the conversation.
17. The completed exercise can be re-accessed.
18. Able to tell jokes.
19. The ability to customize the interface of the chatbot.
20. Include a community of users where they can communicate.
21. Monitor the mood of the user.

B

Design Appendices

B.1. System categorization

Based on the taxonomy of conversational agents in health, proposed by Montenegro et al. [88], the intended chatbot support can fall into the following categories: for the context, the chatbot is a therapy-based system targeted toward PhD students who can be seen as patients (not physicians nor psychology students). The system would assist the PhD candidates in overcoming motivational barriers of performing thought strengthening exercises. As for the dialogue type, the system will be a "dialogue management" type, which means it would be in charge of the dialogue state of the interactions with the chatbot. As the text-based software provides means to alter PhD students' behaviour toward completing CBT homework, it can be categorized as "coaching agents" and "text-based communication" types. The last category in this taxonomy is regarding the architecture, mainly the techniques and the tools used. The chatbot is considered as a semantic-based software that uses Rasa as its development framework.

B.2. Design recommendations

Various mental health softwares (e.g., Woebot and Wysa) integrated some of the recommendations for developing an evidence-based intervention listed in [4]. These recommendations were tested and have yielded promising results. As it is usually not possible to integrate all recommendations, we adopted 11 out of the 16 requirements. Table 8 shows a comparison between our system, Woebot and Wysa regarding integrating the design recommendation.

Furthermore, the chatbot integrates the applicable best practices for conversational agent usability in mental health, which were listed in [26]. These practices are: delays and typing indicators while the user is waiting for an answer, pick up and use the name of the users in the conversations to maintain harmony between the user and the chatbot, include conversations from real human-to-human sessions in the same context as the agent, and inclusion of jokes, gifs and emojis in reacting to situations. The next section details how these practices were included in the chatbot.

Table B.1: A comparison of the design recommendations integrated in the motivation system, Woebot and Wysa

Design Recommendation	**Motivation system**	**Woebot**	**Wysa**
1) CBT based	✓	✓	✓
2) Address both anxiety and low mood	x*	✓	✓
3) Designed for nonclinical populations	✓	✓	✓
4) Automated tailoring	✓	✓	✓
5) Reporting of thoughts, feelings, or behaviors	✓	✓	✓
6) Recommend activities	✓	✓	✓
7) Provision of mental health information	✓	✓	✓
8) Real-time engagement	✓	✓	✓
9) Activities linked to specific reported mood problems	✓	✓	✓
10) Encourage non technology-based activities	x	✓	✓
11) Gamification and intrinsic motivation to engage	✓	✓	✓
12) Log of past use	x	x	x
13) Reminders to engage	x	✓	✓
14) Simple and intuitive interface and interactions	✓	✓	✓
15) Links to crisis support services	✓	✓	✓
16) Experimental trials to establish efficacy	x	✓	✓

* Contradicts with the system focus.

C

Evaluation Appendices

C.1. Technical details

C.1.1. Architecture

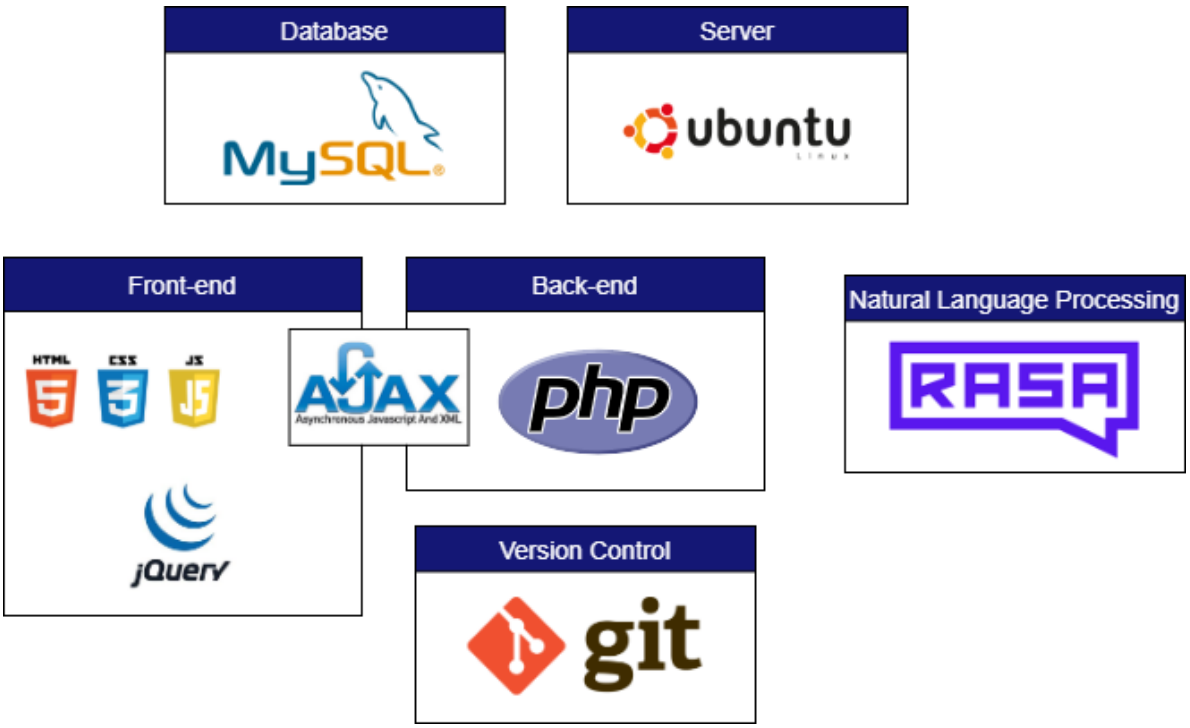


Figure C.1: Software architecture of the prototype.

C.1.2. General dialogue tree

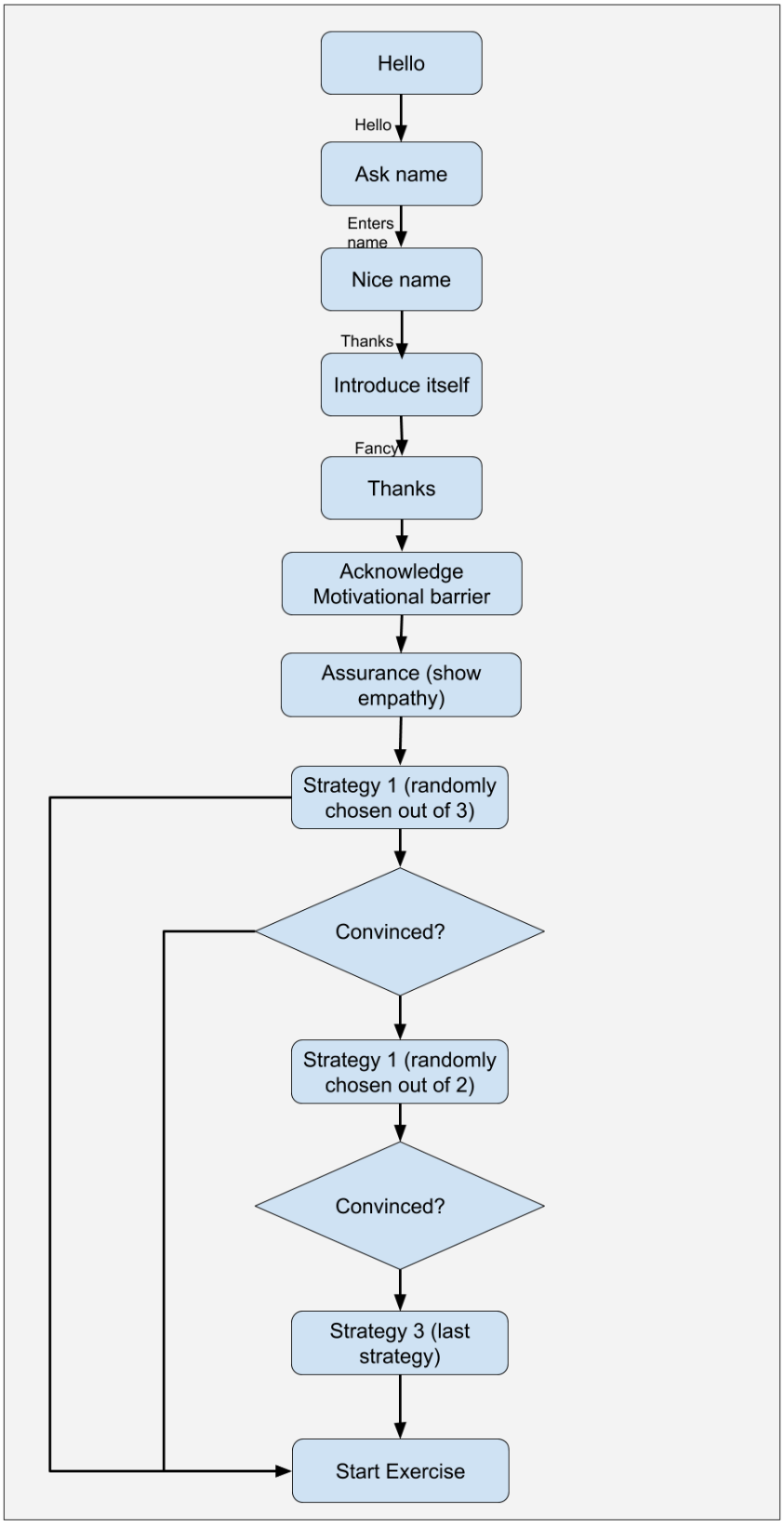


Figure C.2: An overall dialogue tree for the different sessions.

C.2. Experiment

C.2.1. Informed consent form

Thank you for considering to participate in this experiment. The experiment consists of typical task in cognitive therapy. Cognitive therapy aims to change negative, debilitating thinking patterns to more positive ones, as this can lead to improvements in mood. We are conducting this experiment to better understand how people go about changing their negative thoughts.

This study is divided into four parts: (1) instructions and demo video, (2) questionnaire number 1, (3) the task, and (4) questionnaire number 2. In the task, you will be given a description of a situation. Also, we will provide you with a negative thought that causes you to feel bad in that situation and a healthier positive thought. You will be asked to find arguments for and against the negative and positive thoughts. Before you start on the task, there will be a video demo of how to do the task which will be followed by the first questionnaire. After finishing the task, you will be directed to a second questionnaire. The questionnaire completes the study.

During the experiment, you may be asked to enter a name that you make up for this experiment. The entered name may then be used to address you but will not be stored. Any change in the web page (e.g., going to the next page, closing the browser etc..) will result in the name being deleted permanently. In addition to the questionnaire answers, information on your interaction with the website will be stored.

To minimize the risk of data leakage, all the retrieved data (consent form and all the other data mentioned above) will be anonymized by the principal investigator before storing data, and no identifiable data will be registered. The retrieved data will be stored in 4TU Center for Research Data, after the study is terminated.

After filling out the questionnaire, you will be paid $3.50.

Please note that participation is voluntary and before submitting the final questionnaire, you can withdraw your participation at any time and no data will be registered. After submitting the second questionnaire, you will not be able to withdraw your responses anymore.

If you experience issues with your thoughts and feelings as a result of participating in the study, then consider contacting your local mental health helpline:

https://checkpointorg.com/global/

Table C.1: Consent form checklist

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
- I have read and understood the study information dated [19/02/2020]. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	☐	☐
- I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study during the study, without having to give a reason. I also understand that I can no longer withdraw from the study after submitting the questionnaire.	☐	☐
- I understand that taking part in the study involves doing a writing arguments task and a questionnaire completed by me.	☐	☐
Risks associated with participating in the study		
- I understand that taking part in the study involves the following risks: possible data leakage. This risk is mitigated by the researcher by storing questionnaire and task responses in an anonymous format. Access to concrete responses is only given to the researcher and his daily supervisor. No identifiable data will be registered.	☐	☐
Use of the information in the study		
- I understand that the information I provide will be used for scientific publications by the researcher.	☐	☐
- I understand that personal information collected about me that can identify me will be removed from the collected data before storage.	☐	☐
Future use and reuse of the information by others		
- I give permission for the questionnaire responses, information on my interaction with the website and answers from the writing arguments task that I provide to possibly be archived so it can be used for future research and learning.	☐	☐
By checking this box, I agree to take part in this research. ☐		

C.2.2. Strategies text

Not Convinced

It's alright if you do not fully believe that the positive thought can replace the negative thought. After forming the positive thoughts, they may not be as believable as the negative thought. Therefore, the positive thoughts need to be strengthened through writing arguments that support them and weakens the negative thought. As your belief in the positive thought increases, your improved mood will stabilize. Research also has found that doing writing arguments exercise positively predicts successful outcomes in therapy. Kazantzis and others combined the results of 32 studies and found that when people did more arguments writing exercise they had better outcomes. It is also considered as an essential part of therapy by many researchers.

The following story about Tom was told by Kazantzis and others in their book "Using Homework Assignments in Cognitive Behavior Therapy". Tom was an accountant who was having difficulty with work, decision-making and had regrets toward past investments. So, he was afraid to make any decisions now that he might regret later. To change this thought, Tom was asked to do arguments writing exercises, which could have changed his negative thinking. After doing it for some time, Tom started to realize that he might be overreacting and coming up with frightening scenarios in his head, which made him more certain that things were really terrible.

Therapists consider the arguments writing exercise an essential part of cognitive behavioural therapy, which is a therapy that is widely used due to its effectiveness. Therapists often see exercise completion as a prediction of whether the therapy will work. Also, Aaron Beck, the founder of cognitive behavioural therapy and one of the "five most influential psychotherapists of all time," considers the argument writing exercise an important part of cognitive behavioural therapy.

Low Self-efficacy

You indicated that you have some worries regarding your ability to do the exercise. Your reaction regarding your ability to write arguments is normal at this stage. It is like a part of the game. In the beginning, it may be difficult, but when you are past the most difficult period it only helps you. You are a go-getter and you will do very well!. You're well on your way! You can do it.

The following story about Tom was told by Kazantzis and others in their book "Using Homework Assignments in Cognitive Behavior Therapy". Tom was an accountant who was having difficulty with work, decision-making and had regrets toward past investments. So, he was afraid to make any decision now that he might regret later. To change this thought, Tom was asked to do arguments writing exercises, but eventually, he started to avoid doing them. Tom thought he was not able to do the exercises and felt bad about himself. So, he broke down the exercise into smaller tasks to make it easier to do. After writing arguments for some time, Tom felt better because he found the writing arguments exercise became easier the more he did it.

Psychologists consider the arguments writing exercise an essential part of cognitive behavioural therapy, which is a therapy that is widely used due to its effectiveness. Therapists said that people often struggle with doing the writing arguments exercise, usually at the beginning stages. However, therapists also reported the more the individual does this exercise, the easier it gets. They also say it's ok if the exercise is not done all the way, but it's better to be done nonetheless.

Perfectionism

You indicated that you have some worries regarding your high expectation of your performance for doing the exercise. It's common while writing arguments to try and do everything perfectly. Remember that you should not strive for perfection while doing the exercise. The exercises are not meant to be answered perfectly. Learning to write arguments is a skill, like learning the computer. You'll get better with practice.

The following story about Andrew was told by Kazantzis and others in their book "Using Homework Assignments in Cognitive Behavior Therapy". Andrew was on a holiday in Turkey when England beat Turkey at a football match and he wore his England shirt the next day. In a restaurant, he noticed Turkish people giving him hostile looks, and he became increasingly worried that he caused offense and may be harmed during his stay. To change this thought, Andrew was asked to do arguments writing exercises, but he feared that he would make small mistakes while writing the arguments. So, he tried to do the exercises without worrying about writing everything perfectly. After writing arguments for some time, Andrew felt better as he noticed an improvement in his mood despite making smaller mistakes.

Therapists consider the arguments writing exercise an essential part of cognitive behavioural therapy, which is a therapy that is widely used due to its effectiveness. The exercise completion is often seen by thera-

pists as a prediction of whether the therapy will work. Therapists say it's ok if the exercise is not done perfectly right, but it's better to be done nonetheless.

C.2.3. Perceived Usefulness questionnaire

We are very interested in your opinions here, so, there are no right or wrong answers.

The following questions are about your personal perceived usefulness for the arguments writing exercise. Imagine that you have a negative thinking pattern in the future, and that you are asked to do the exercise at home in order to change it:

Table C.2: Perceived usefulness questionnaire

	Strongly Disagree	**Disagree**	**Neither agree nor disagree**	**Agree**	**Strongly agree**
I believe writing arguments would be of some value to me.	☐	☐	☐	☐	☐
I think that doing writing arguments task would be useful for changing my negative thoughts	☐	☐	☐	☐	☐
I would be willing to do the writing arguments task because it has some value to me.	☐	☐	☐	☐	☐
I believe doing this writing arguments task could be beneficial to me.	☐	☐	☐	☐	☐
I think writing arguments would be an important activity to me.	☐	☐	☐	☐	☐
Overall, I would find writing arguments to be advantageous for changing my negative thoughts.	☐	☐	☐	☐	☐

C.2.4. Self-efficacy questionnaire

Imagine that you are a person who has a negative thought which you want to change, so you are asked to complete an arguments writing exercise.

Rank the following situations from 0 (highly certain cannot do) to 10 (highly certain can do) regarding how certain you are that you can complete the arguments writing exercise:

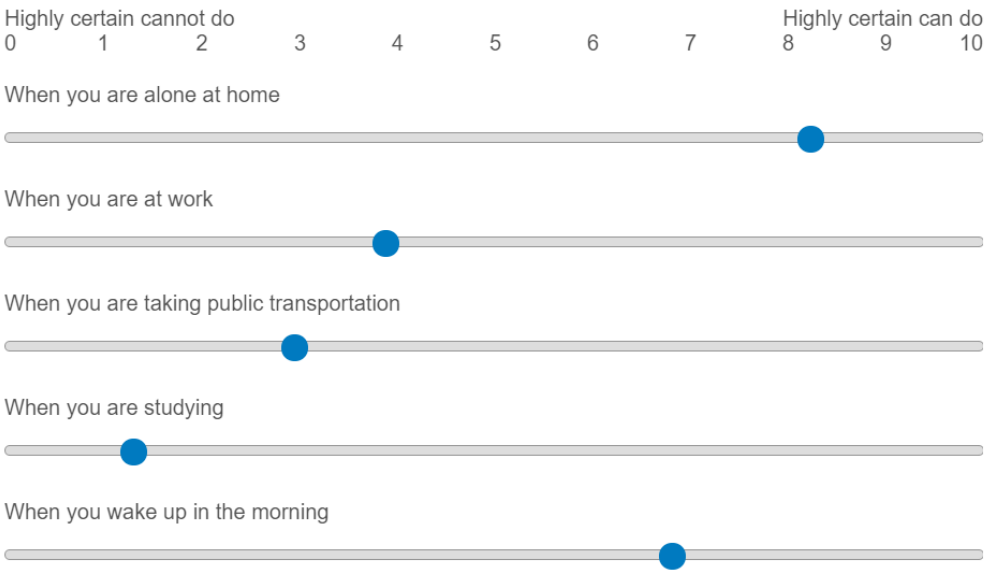


Figure C.3: Self-efficacy questionnaire, taken from the survey.

C.2.5. Other questionnaires

Enjoyment question

Now, we want to compare how you like the arguments writing exercise to picture-labeling tasks. Picture-labeling tasks refer to writing down what you can see in a picture (for example, the first picture is a pencil and

then you write "pencil", the second picture is an elephant, so you write "elephant", etc..)

Imagine that, as an Amazon Mechanical Turk worker, you have to spread 20 minutes in doing picture-labeling tasks and arguments writing exercises. How would you distribute your 20 minutes?

Figure C.4: Enjoyment question, taken from the survey.

Enjoyment questionnaire

We tried to help you overcome the difficulty that you rated the highest (your interaction with the chatbot

How much did you enjoy the chatbot part of the experiment?

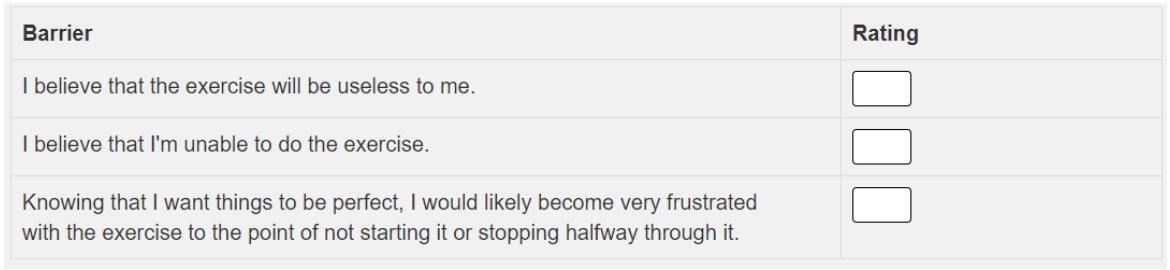
Table C.3: Enjoyment questionnaire

	Strongly disagree	**Disagree**	**Neither agree nor disagree**	**Agree**	**Strongly agree**
I enjoyed that part very much.	☐	☐	☐	☐	☐
That part was fun to do.	☐	☐	☐	☐	☐
I thought that part was a boring activity.	☐	☐	☐	☐	☐
That part did not hold my attention at all.	☐	☐	☐	☐	☐
I would describe that part as very interesting.	☐	☐	☐	☐	☐
I thought that part was quite enjoyable.	☐	☐	☐	☐	☐
While I was doing that part, I was thinking about how much I enjoyed it.	☐	☐	☐	☐	☐

Barriers rating

Again, imagine that you have a negative thinking pattern in the future and are asked to do the arguments writing exercise at home in order to change it.

What would be the reason that you may find it difficult to do the arguments writing exercise? Rate the following statements from 0 (not applicable) to 100 (very applicable):



Barrier	Rating
I believe that the exercise will be useless to me.	
I believe that I'm unable to do the exercise.	
Knowing that I want things to be perfect, I would likely become very frustrated with the exercise to the point of not starting it or stopping halfway through it.	

Figure C.5: Barrier ratings, taken from the survey.