Master`s Thesis

July 2022

The Role of Personalization and Engagement in Digital Health Interventions for Depression

Kai Vögeding

s2084031

Health Psychology and Technology

Faculty Behavioural, Management and Social Sciences

1st Supervisor: Dr. A. van Dongen

2nd Supervisor: Dr. S.M. Kelders

UNIVERSITY OF TWENTE.

Contents

Abstract ... 1
1. Background ... 2
2. Method ... 7
2.1 Design ... 7
2.2 Participants ... 7
2.3 Materials ... 8
2.3.1 The Incredible Intervention Machine (TIIM App) ... 8
2.3.2 TWEETS questionnaire ... 11
2.3.3 PHQ-9 questionnaire ... 13
2.3.4 GAD-7 questionnaire ... 13
2.3.5 MHC-SF questionnaire ... 13
2.4 Procedure ... 13
2.4.1 Personalization Procedure ... 14
2.5 Data analysis ... 15
3. Results ... 17
3.1 Descriptive Statistics ... 17
3.2 Intervention effectiveness ... 19
3.2.1 Research question 1 ... 19
3.2.2 Research question 2 ... 20
3.3 Engagement as mediation ... 20
3.3.1 Research question 3 ... 20
4. Discussion ... 21
4.1 Main findings ... 21
4.2 Research questions 1 and 2 ... 22
4.2.1 Personalization of the intervention ... 22
4.2.2 Factors of the Intervention (TIIM App) ... 23
4.3 Research question three ... 25
4.3.1 Mediation Model ... 25
4.3.2 Negative correlation path b ... 26
4.4 Strengths and limitations ... 27
5. Conclusion ... 28
6. References ... 30
7. Appendices ... 40

Abstract

Background. Due to different reasons, people with mental problems encounter difficulties in seeking sufficient help. The healthcare system cannot serve the demand for mental health treatment. Compared to traditional treatment, technologies like Digital Health Interventions (DHIs) have advantages that may help to regulate these difficulties. Research has shown that DHIs are effective in treating health problems. However, a low level of engagement was found to be the main barrier to reaching the desired effects. This study aimed to investigate the influence of personalization of DHIs on the engagement outcome scores as well as the depression outcome scores. A mediation model was used to determine the mediating role of engagement in the relationship between personalization and depression outcome scores.

Methods. The final sample size consisted of 176 participants who participated in a 14-day mobile (digital) health intervention (TIIM App) with daily tasks based on evidence-based treatment approaches such as positive psychology. Participants were randomly assigned to the experimental or control group. A personalization approach was used to determine the best matching version of the intervention for the participants within the experimental group. Personalization was conducted through three interventions and technological factors (ITFS) namely *content, feedback,* and *design.* Independent sample t-tests and repeated measure ANOVA were used to determine the main effects of personalization on both engagement and depression. A mediation analysis was performed to test whether engagement mediated the relation between personalization and depression.

Results. A main effect for time on change in depression scores was found (Wilks lambda = .908 $F(1,173)$ =17.56, p < .001). No significant differences between the conditions experimental group (with personalization) and the control group (no personalization) were found in the outcome measures of depression and engagement. In addition, no mediating effect of engagement has been found for the relationship between personalization and engagement. Within the mediation analysis, a negatively significant effect was found for the relationship between engagement and depression (B = -.23, SE= .04, p= <.01).

Conclusion. The non-significant results found in this study were quite surprising and not expected. The negative relation between engagement and depression scores could be interpreted in two different ways. However, these results were expected and confirmed the hypothesis. Future research is recommended to improve the overall design and quality of the intervention as well as the personalization procedure. Moreover, besides personalization, other factors should be considered regarding their influence on engagement and personalization.

1. Background

Based on epidemiological studies it can be stated that depression is one of the most common mental disorders all over the world (Kessler & Bromet, 2013). According to Wittchen and Jacobi (2005), since 1990 depression has been the most burdensome disease in Europe ahead of other diseases like coronary heart disease or diabetes mellitus. In Germany, around 28% (17,8 million people) of the population suffer from mental disorders yearly (Jacobi et al., 2014). It must be mentioned that these numbers have increased due to the COVID-19 pandemic. A Dutch study published in 2021 found that 30% of the participants reported a higher feeling of loneliness, stress, and sadness during the corona pandemic than before (National Institute for Public Health and the Environment, 2020). In spring 2021, the number of young German patients was ⅓ higher compared to the pre-pandemic period (Mangiapane et al., 2020). Based on research by the European Commission (2018), the total costs related to mental disorders in Germany including indirect and direct costs are 147 billion euros each year.

However, many people that have to deal with mental problems like depression do not receive adequate treatment. Only 18,9% of the affected individuals in Germany get in contact with health care professionals in order to treat their psychological problems (Mack et al., 2014). Moreover, the time span between the initial contact and the start of the therapy consists on average of 5 months (Bundespsychotherapeutenkammer, 2018). Aside from the high treatment costs that an individual might have to face in the conventional way of treatment, other different disadvantages and barriers might hinder people with mental disorders, like depression, to receive treatment. Certain social and cultural groups have problems accessing treatment by professionals due to financial or geographical reasons (Glied et al., 2003; Mennis et al., 2012; O'connor et al., 2016; Radez et al., 2021). Even nowadays people with mental disorders often experience stigmatization based on their disorder (Malla et al., 2015; Tyerman et al., 2021). It was found that this can especially create a barrier t treatment for people with low self-confidence. They hesitate to receive professional treatment based on the fear of not being accepted by the psychologist and society (Andrade et al., 2014). However, it has to be mentioned that people with a healthy level of self-confidence might experience these fears as well. Another problem might be that people are often not aware of the importance of treatment and have the perception that it is sufficient to treat themselves (Saxena et al., 2007). To summarize, it can be stated that the capacities provided by the health care system are not sufficient to offer adequate treatment for all people in need.

With the growing digitalization, the professionalization of eHealth put forth an alternative treatment to the traditional methods (Du et al., 2020). eHealth, also called digital health, is used as an approach to treat different physiological and psychological disorders. This paper will focus on technologies that help to treat mental disorders like depression. This approach is called eMental Health (eMH). Lal (2019) defines eMental Health as the "use of the Internet and related technologies to deliver or enhance mental health information and services". Hereby, different methods are used to deliver or enhance mental health information and services. One of these methods is digital health interventions (DHI). Blandford et al. (2018) describe digital health interventions "as interventions designed to improve health that are delivered on a digital platform". Digital platforms can be for example mobile devices, technological sensors, or virtual reality.

Digital health interventions contain many benefits compared to traditional ways of treatment. Therefore, mental health care services provided through technological methods have become even more interesting for the healthcare system. First, compared to traditional treatment, DHIs can be associated with lower service costs (Andrews et al., 2010). For instance, DHIs can reach large populational groups through one digital platform like an app (Hedman et al., 2012). DHIs, give the chance to support and treat people with mental health problems at different stages of the disease. For example, they can help to treat the first symptoms of depression at an early stage of the disorder. . Early identification can prevent a severe course of depression and high treatment costs (Hollis et al., 2015). Consequently, DHIs can also be used for the general management of mental disorders as well as the evaluation of treatment processes (Fortuna et al., 2015). In addition, it can be argued that DHIs offer easier accessibility for people who suffer from mental problems for the first time. DHIs are often not dependent on healthcare professionals or timeslots. They overcome the barrier of doubts and fear of contacting a health care professional as well as long waiting periods (Andersson, & Cuijpers, 2008; Taylor-Rodgers, & Batterham, 2014). Compared to face-to-face treatment DHIs are flexible in treatment options such as time, place, or language (Handley et al., 2014; Meadows et al., 2015). As a result, the content of DHIs can be adapted to the needs and circumstances of the people that use these interventions. Lastly, Klein et al. (2016) and Meyer et al. (2014) found that DHIs can promote the self-guided treatment of depressive symptoms. Individuals can track their state of disease and make decisions based on the information provided by the DHIs.

As mentioned above, DHIs have many advantages compared to the traditional way of treatment. Research shows that DHIs can effectively reduce symptoms of depression. For example, different studies found that computerized cognitive behavioural therapy (cCBT) minimizes symptoms of depression significantly (Pennant et al., 2015; Richardson et al., 2010). Topooco et al. (2018) used a chat and internet-based cognitive behavioural therapy in which they treated depressive symptoms of adolescents. Participants with mild and moderate depression were asked to do eight different skill-based modules which included reading material, educational videos, or interactive tasks. The results showed that after eight weeks the depressive symptoms decreased by 50% within 42% of the experimental group. A meta-analysis from Firth et al. (2017) indicated that smartphone-based mental health interventions promote the self-management skills of people suffering from depressive symptoms. Consequently, it can be argued that DHIs have the same effectiveness in reducing depressive symptoms as face-to-face treatment (Erbe et al., 2017).

As mentioned in the previous paragraphs DHIs consists of many benefits and chances to improve the well-being of individuals that suffer from depressive symptoms. However, they also have some limitations and points to consider. Numerous barriers prevent people from participating in DHIs such as being too busy, feeling incapable of using the technology or disliking its impersonal nature (Gorst et al., 2014; Sanders et al., 2012). A lot of people that intend to use DHIs stop using them shortly after downloading them. This has the consequence that positive effects cannot be reached or do not last for a long period of time (Molloy & Anderson, 2021). An investigation from 2016 found that within the broad sector of mobile apps around 70% of the app users stop using an app within the first week after downloading (Sigg et al., 2016). According to Kelders et al. (2020a), low user engagement is an important factor in why DHIs do not reach their full potential. The concept of engagement in DHIs is barely defined. Perski et al. (2017) defined engagement in the context of DHIs "as the extent of usage like amount, frequency or duration and the subjective experience characterized by attention, interest, and affect". Most of the studies define engagement as a behavioural construct that focuses on the frequency of use and the routine of using DHI in daily life (Kelders et al., 2020b). Nevertheless, Kelders et al. (2020b) expanded the definition of engagement in DHIs as a concept that includes behavioural as well as cognitive and affective components. Studies have shown that engagement has an impact on the effectiveness of DHIs (Yardley et al., 2016). To date, the concept of engagement in DHIs is barely investigated. Recent studies recommended determining the factors that influence engagement in DHIs that aim to improve the well-being of people with depressive symptoms (Aref-Adib et al., 2019; Kelders et al., 2020a; Perski et al., 2017).

Nowadays researchers set a specific focus on the personalization of DHIs to increase user engagement and its effectiveness. Sebri & Savioni (2020), defined the personalization of medicine and healthcare “as the medical approach that uses the specific biological characteristics, environment, needs, and lifestyle of an individual to create ad hoc therapy, including drugs, dosages, and other possible remedies”. Using this definition in the context of DHIs it can be argued that the design and content of an intervention should be tailored toward the personal needs of an individual to reach the most effective treatment for this person. For instance, studies from the field of nutrition and weight-loss management found that the personalization of weight-loss management interventions was linked to higher levels of engagement which in consequence led to better results in weight loss of the participants (Dennison et al., 2014; Tang et al., 2015). Personalization was implemented in form of individualized feedback and encouragement. Within the experimental group, the content of the intervention was tailored towards the personal preferences of the participants which resulted in higher satisfaction with the design of the intervention and had a positive effect on the level of engagement. On the one hand, research in the field of weight-loss management stated that the personalization of DHIs is an important factor to increase engagement, which in turn increases the effectiveness of DHIs. On the other hand, these studies formulate the need for further investigations to find ways how the factor of personalization in DHIs can be effectively implemented (Saperstein et al., 2007; Watson et al., 2015). Within the field of eMental Health little research exists on if and how personalization influences engagement in DHIs (Kelders, 2019). However, personalization of DHIs is needed to overcome the barriers of generalized DHIs which might be related to the high dropout rates. Different studies stated that personalized contents and designs of DHIs may lead to positive associations with the DHI and increase engagement. Consequently, people feel more motivated to use the intervention regularly (Kelders, 2019; Perski et al., 2016). To summarize, it can be stated that the personalization of DHIs seems to be an effective and efficient approach to promoting the engagement in DHIs of users who aim to treat their symptoms of depression.

The concept of personalization to promote engagement in DHIs can be implemented in many ways. Kelders (2019) used three different elements to ensure personalization in DHIs. These elements were also used in this study. First, the intervention can be more personalized if the content is in line with the personal beliefs and values of the user (Hayland & Whalley, 2008; Whalley & Hyland 2009). Furthermore, the opportunity to choose from different types of feedback channels makes it possible to create a personal version of the intervention and get potentially more engaged with it (Groeneveld 2020; Talbot, 2012). Lastly, some individuals enjoy it when certain applications consist of gamification elements. The element of gamification can increase the engagement of users in DHIs (Hamari et al., 2014). According to

Kelders (2019), it is important to do further research about different options that personalize DHIs and enhance the engagement of users which may lead to higher effectiveness of DHIs. This study aimed to investigate the influence of personalization of DHIs on the engagement outcome scores as well as the depression outcome scores. The following research questions were formulated:

RQ 1: *Does the personalization of a 2-week Digital Health Intervention for depression and wellbeing directly influence the effectiveness of digital health interventions?*

- *Do people in the experimental group with a personalized 2-week Digital Health Intervention show a lower level of depression than people out of the control group after 21 days of intervention?*

RQ 2: *Does the personalization of a 2-week Digital Health Intervention for depression and wellbeing directly influence the participants' level of engagement?*

- *Do people in the experimental group with a personalized 2-week Digital Health Intervention show a higher level of engagement than people out of the control group after 21 days of intervention?*

Based on previous research findings it can be suggested to investigate if engagement mediates the relationship between the factor of personalization and the effectiveness of DHIs. Therefore, the following research question will be investigated:

RQ 3: *Does engagement mediate the relationship between personalization and the overall effectiveness of a 2-week Digital Health Intervention for depression and wellbeing?*

- *Does the personalization of a 2-week Digital Health Intervention increase the participants' level of engagement which in turn increases the effectiveness of the overall intervention in reducing the symptoms of depression?*

2. Method

2.1 Design

This research is part of a larger study currently executed by Kelders at the University of Twente (UT). The main purpose of the larger research project was to develop a valid personalization approach for eMental Health. In this study, an experimental randomized controlled trial was used to determine if a personalization approach within the experimental group leads to a higher engagement with the DHI and in turn to a higher reduction in depressive symptoms compared to the control group which did not receive the personalization approach. The overarching study used three different intervention and technological factors (ITFs) namely content, feedback, and the design of the intervention. These ITFs were adopted and used in the current study. The participants were asked to work with the intervention for 21 days. Investigations found that smartphone-based interventions showed significant results in the reduction of depressive symptoms after 14 days (Fitzpatrick et al., 2017; Lukas et al., 2021). Moreover, they had to fill out different questionnaires at three different time periods. At baseline (T0), post (T1), and follow-up (T2) (Appendix A). Quantitative data analysis was deployed to get insights in the collected data. This study was approved by the Ethics Committee of the faculty of Behavioural, Management and Social Science (BMS) at the University of Twente (approval code: 220083).

2.2 Participants

The sample of the study were recruited through the SONA system. Students from the University of Twente had the possibility to enroll in the study through the SONA system platform of the University. They were rewarded with so-called ‘’SONA credits’’ which are obligatory for a successful graduation process at the University of Twente. Participants were able to participate when they were interested in the study, had access to a smartphone, spoke English, and were older than 18 years old. A power analysis (G*Power) (1-β=0.80; alpha=0.05; expected Cohen’s d=0.40) was used to calculate the needed sample size. The expected dropout rate was set at 35% which resulted in calculated recruitment of n=306. The power analysis was taken from the grant application form (research proposal by Kelders, 2017). Participants were excluded from the sample when they had a low score of depressive symptoms indicated by the results of the Patients Health Questionnaire-9 (PHQ-9) or because they had low anxiety complaints indicated by the results of the Generalized Anxiety Disorder Questionnaire-7 (GAD-7). The cutoff point was all total scores below five. In addition, participants had to be excluded because of the following reasons: they did not fill out the baseline survey, they did not register for the TIIM

App, or they did not fill out the post-survey. Participants that stopped participating in the TIIM App intervention were also considered in the results. These participants were not excluded from the analysis because it can be hypothesized that either the personalization of the DHI or the randomly assigned version of the DHI was not engaging enough to continue with the intervention.

2.3 Materials

2.3.1 The Incredible Intervention Machine (TIIM App)

The intervention of the study was executed through the TIIM app. The TIIM app was provided by the University of Twente. It is a tool that gives students and employees from the UT the opportunity to design and monitor different kinds of digital interventions. The TIIM app was used to create the intervention used in the current study. The design process and implementation of the intervention to improve well-being were part of the larger eMental Health research project executed by Stewing (2021) and Wehrmeyer (2021). The leading researchers of the larger research project, as well as the BMS Lab of the University of Twente, gave permission to use the intervention for the aim of the current study. The intervention consisted of 27 different versions compounded by the combination of three different ITFs (content, feedback, and design). Each ITF consisted of three different options (3x3x3 research design, Appendix B). The researcher was able to manage the different versions of the intervention on the software of the TIIM.

The terms *factor, options*, and *version* will be explained in the following paragraph. These terms will be used throughout the whole article. The term *factor* describes the three different intervention and technological factors *content, feedback,* and *design*. The term *option* stands for the different options within one factor. For example, the factor *content* consists of the options positive psychology, meaning interventions, and cognitive behavioural therapy. The term *version* describes the combination of different options between different factors. For example, the combination of the options positive psychology (*content*), plain written text (*feedback*), and no gamification elements (*design*) was one of the 27 versions.

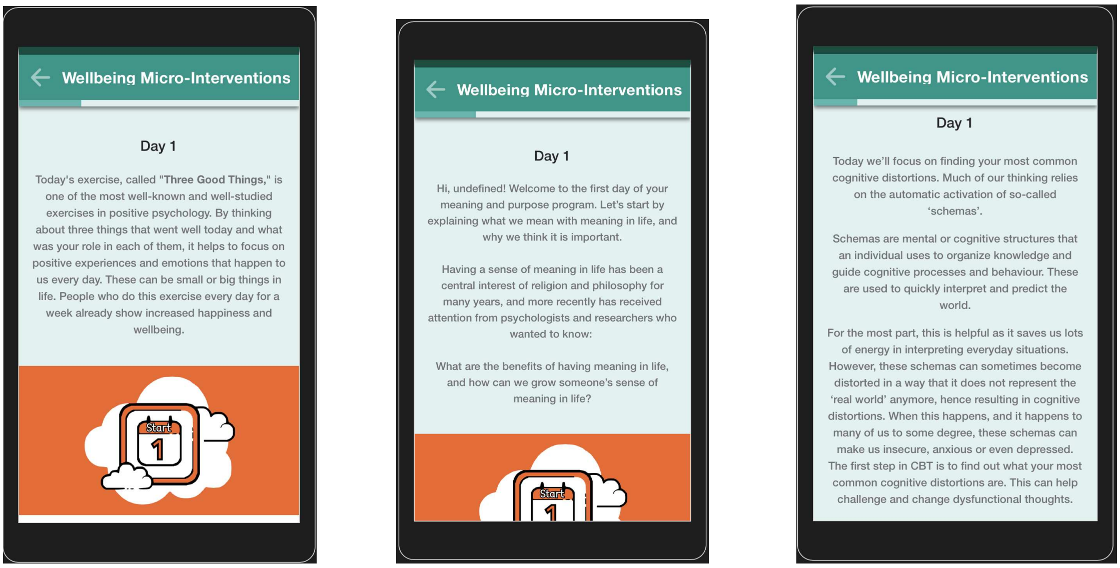
2.3.1.1 Factor content

The different options of the three different ITFs will be explained in the following paragraphs. First, the factor *content* consisted of three different psychological theories which aimed to give the theoretical background for the daily tasks that had to be done within the intervention. Based on evidence-based literature three psychological approaches have been used to implement the different exercises: Cognitive Behavioural Therapy (CBT), Meaning Intervention (MI), and

Positive Psychology (PP) (Kelders, 2019; van Agteren et al., 2021). For instance, one exercise within the approach of positive psychology was that participants had to write down good memories from the past. This would help them to learn what it means to consciously think about positive emotions and promote these (Burton & King, 2004). The content of the exercises was different between the psychological approaches. However, they all had the aim to promote well-being and minimize the symptoms of depression. An example of the different theories and how they approach the exercise of day one can be seen in Figure 2.

Figure 2

Examples of different exercises of the three different content options



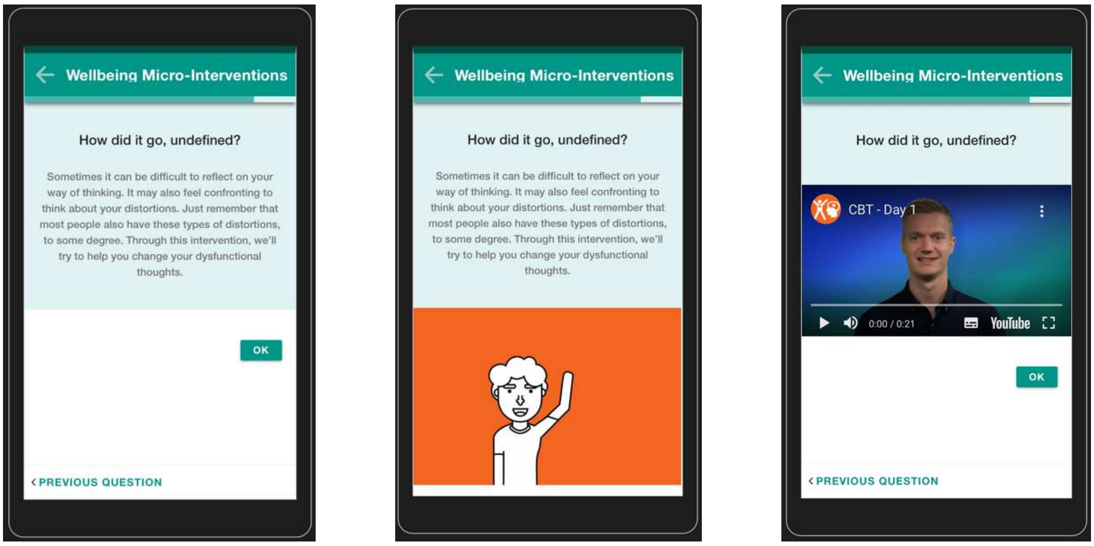
Note. From left to right: the content of the interventions in the form of the following psychological approaches (1) Positive Psychology, (2) Meaning Intervention, (3) Cognitive Behavioural Therapy.

2.3.1.2 Factor feedback

The factor *feedback* was used to give the participant some feedback after they completed the daily task. The feedback was delivered in three different ways. First, a plain text was shown after the exercise. No additional illustrations were part of the feedback. As a second option participants received the feedback in form of a virtual agent. Figure 3 shows how the text was the same as in version one except for the difference that option 2 also showed a virtual agent, in this case a cartoon person, that "provided" the feedback.. The third option consisted of a pre-recorded video without any text information. A human counselor gave the participant feedback in spoken words. It must be mentioned that the content of the feedback message was the same in all three different options.

Figure 3

Examples of different ways to give feedback in the intervention



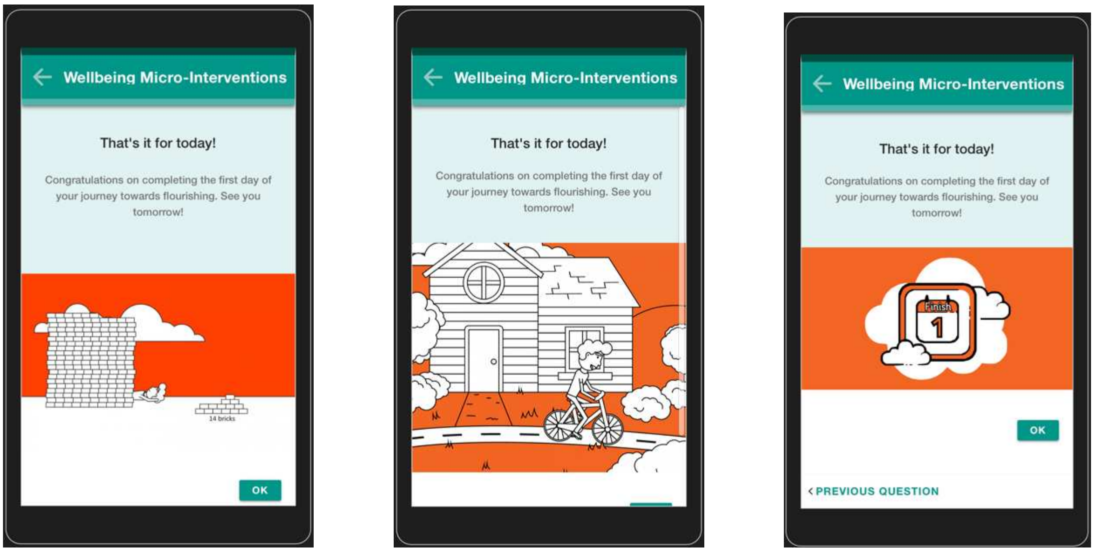
Note. From left to right: provided feedback of the exercises as (1) a plain written text, (2) a written text provided by a virtual agent, (3) spoken words by a human counselor in a pre-recorded video.

2.3.1.3 Factor design

The third factor was about the *design* of the different intervention options. This factor was also distinguished into three different designs. The first design option of the intervention consisted of gamification elements that were competitive. During the progress of doing the exercises, participants saw a stack of bricks moving from left to right. The second option included gamification elements that were noncompetitive. In this option, the intervention progress was illustrated by a virtual avatar that cycled to different places (Figure 4). This had the purpose to create a storyline for the participant. Lastly, the third version was not gamified. The participant saw a certain day of the intervention. The participant was shown a date on a calendar that related to the day of the intervention as shown in figure 4.

Figure 4

Examples of different intervention designs



Note. From left to right: intervention design as (1) competitive gamification, (2) non-competitive gamification, and (3) no gamification elements.

2.3.2 TWEETS questionnaire

The Twente Engagement with eHealth Technologies Scale (TWEETS) questionnaire was one of two questionnaires used in this study. The TWEETS questionnaire was developed to measure the level of engagement in eHealth technologies like the intervention of this study (Kelders et al., 2020b). It consists of nine items in total measured on a 5-point Likert scale that varies between strongly disagree and strongly agree. The total engagement scores range between 9 (not engaged) and 45 (highly engaged) (Kelders et al., 2020b). The questionnaire determined the level of engagement in three different areas namely behavioural engagement, cognitive engagement, and affective engagement (Kelders & Kip, 2019). Each of them consists of three items. For the aim of the current study, the items of the questionnaire were adjusted to the content and app specific elements (Table 1). The level of engagement was assessed at baseline- and post-intervention. The TWEETS questionnaire at the baseline (Table 1) was used during the personalization procedure in order to personalize the intervention of the experimental group (paragraph 2.4.1). The TWEETS questionnaire at post-measurement was used to obtain the participants level of engagement 21 days after they have started with the DHI (Appendix C).

Table 1

Adjusted TWEETS items for baseline questionnaire assessing engagement scores regarding content and app-specific elements

Item	Content-specific TWEETS	App-specific elements TWEETS
1	Using an app with this *content* can become part of my daily routine.	Using this *version of the app* can become part of my daily routine.
2	The *content* of this app is easy to use.	This *version of the app* is easy to use.
3	I will be able to use an app with this *content* as often as needed to improve my well-being.	I will be able to use this *version of the app* as often as needed to increase my well-being.
4	An app with this *content* will make it easier for me to work on increasing my well-being.	This *version of the app* will make it easier for me to work on increasing my well-being.
5	This *content* motivates me to increase my well-being.	This *version of the app* motivates me to increase my well-being.
6	This *content* will help me to get more insight into my well-being.	This *version of the app* will help me to get more insight into my well-being.
7	I will enjoy using an app with this *content*.	I will enjoy using this *version of the app*.
8	I will enjoy seeing the progress I make by using an app with this *content*.	I will enjoy seeing the progress I make in this *version of the app*.
9	An app with this *content* will fit me as a person.	This *version of the app* will fit me as a person.

Note. Content-specific TWEETS were used for the factor *content* including the options Positive Psychology, Meaning Intervention, and Cognitive Behavioural Therapy. App-specific elements TWEETS were used for the factors *feedback* and *design*. The factor *feedback* included the options feedback provided as a plain written text, a written text provided by a virtual agent, and spoken words by a human counselor in a pre-recorded video. The factor *design* included the options competitive gamification, non-competitive gamification, and no gamification elements.

2.3.3 PHQ-9 questionnaire

The Persons Health Questionnaire (PHQ-9) is a measurement tool that helps to determine depressive symptoms (Appendix D). The PHQ-9 consists of 9 items which are based on the participants' self-administration. Kroenke et al. (2001) explain that every item of the questionnaire is related to one of the nine DSM-IV criteria. DSM-IV refers "to a significant behavioural or psychological syndrome or pattern that occurs in an individual" (Stein et al., 2010). Participants can choose between four different answer options on a 4-point Likert scale ranging from 0 (not at all) to 3 (nearly every day). The total scores varied between 0 and 27 and can be categorized into five different categories namely: Minimal Depression (0-4), Mild Depression (5-9), Moderate Depression (10-14), Moderately Severe Depression (15-19), Severe Depression (20-27) (Kroenke et al., 2001). The cut-off point for this study was everything below five where it could be argued that participants had minimal symptoms of depression. The level of depressive symptoms was measured pre- and post-intervention.

2.3.4 GAD-7 questionnaire

The Generalized Anxiety Disorder (Gad-7) is a self-report questionnaire that aims to assess a person's anxiety level of the last two weeks (Appendix E). The seven items were measured on a 4-point Likert scale ranging from 0 (not at all) to 3 (nearly every day). The total scores ranged from 0 to 21. The cutoff point for this study was five (Williams, 2014). The GAD-7 was included at baseline, post and follow-up. The data obtained from the GAD-7 was not considered and used in the data analysis of this study. This study aimed to investigate the level of depression of the participants and not the level of anxiety. However, this data will be used in the larger research project executed by Kelders at the University of Twente.

2.3.5 MHC-SF questionnaire

The Mental Health Continuum-Short Form (MHC-SF) questionnaire aimed to assess the dimensions of emotional well-being, psychological well-being and social well-being. It consists of 14 items which were measured on a 6-point Likert scale ranging from 0 to 5 (Keyes, 2014). Also, for this questionnaire, the data obtained from the MHC-SF were not considered and used in the data analysis of this study. This data will be also used in the larger research project.

2.4 Procedure

The data collection took part in the period between February and May 2022. The study was promoted through the SONA system. Interested students had the opportunity to receive information about the main purpose of the study, some general information (e.g., theoretical background and content), and the procedure of the study. Students that registered for the study could assess the qualtrics baseline questionnaire through a link provided on the SONAsystem. Before they were able to start with answering the baseline survey they had to read and sign the informed consent of the study. The baseline questionnaire contained questions about the demographics of the participants. Moreover, it included the PHQ-9, GAD-7, MHC-SF, as well as the adjusted TWEETS questionnaire for baseline (Table 1). After filling out the baseline questionnaire participants were checked against inclusion and exclusion criteria. At the end of the baseline questionnaire the remaining participants were asked to download and register for the TIIM App. In addition, the participants received the instructions via email as well. All participants that successfully registered for the TIIM App were randomly allocated to an experimental group or a control group. Participants from the experimental group received the best version of the intervention based on their preferences indicated in the TWEETS baseline questionnaire. Participants out of the control group were randomly assigned to one of the 27 intervention versions. After this process participants were asked to do the daily tasks of their intervention. Ideally, participants worked on the intervention for 14 days straight. However, it was also the case that some participants needed more days to do all the 14 exercises of the intervention. For this reason, they had 21 days to work through the intervention. Participants' progress was regularly checked by the researcher. When inactivity was recognized, they were kindly asked through email to start with the intervention after three and seven days after registration. After 21 days all participants that registered and at least started with the intervention received an email including a Qualtrics link in which they were asked to fill out the post questionnaire which included the PHQ-9, GAD-7, MHC-SF, and the TWEETS questionnaire. All participants that filled out the post questionnaire received another email eight weeks after baseline in which they were asked to fill out the follow-up questionnaire. Participants were asked to create an individual identity code which had to be entered before every response. This helped to match the different responses to one participant. The data obtained in the follow-up survey was not used in this study and was only important for the larger research project by Kelders.

2.4.1 Personalization Procedure

In this study, the TWEETS questionnaire was used to access the level of engagement of the participants. The answers of the participants were used to determine the personal preferences of every participant. The sum scores of the TWEETS questionnaire for each option from each factor were calculated and compared to each other. For example, the sum scores for the options positive psychology, meaning intervention, and cognitive behavioural therapy of the factor *content* were calculated and compared with each other. The option with the highest score on the TWEETS scale was considered to be the most liked one by the participant. The same was done for the options from the other two factors *feedback* and *design*. If participants scored equally between two or three option of one factor they received one option randomly. This was done with the tool random.org. Based on these preferences participants from the experimental group were assigned to the most fitting intervention out of the 27 different versions of the intervention. Participants from the control group were randomly allocated to one of the 27 intervention.

The preferences were assessed as follows. Participants were asked to indicate their level of engagement for each option out of every factor *content, feedback,* and *design*. For example, in the baseline questionnaire participants saw three different templates for each option (CBT, MI, and PP) from the factor *content*. The templates included a description of the belonging theory and an example exercise of the first day of the intervention (Figure 2). For example, the approach of positive psychology in the baseline questionnaire was explained as follows: *"Positive Psychology is a scientific approach to studying human thoughts, feelings, and behavior, with a focus on strengths instead of weaknesses, building the good in life instead of repairing the bad, and taking the lives of average people up to "great" instead of focusing solely on moving those who are struggling up to "normal".* Based on these templates participants had to rate their level of engagement for each category from the factor *content* assessed by the items of the TWEETS questionnaire. The same was done for the factors of *feedback* and *design*. The factor *feedback* was distinguished into plain written text, a written text provided by a virtual agent, and spoken words by a human counselor in a pre-recorded video (Figure 3). And the factor *design* was distinguished into competitive gamification, non-competitive gamification, and no gamification elements (Figure 4).

2.5 Data analysis

For the purpose of the larger research project executed by Kelders data was collected at different time points. Pre-, post-, and follow up- surveys as well as engagement measures on the first, third, and seventh day of the intervention within the TIIM app. For the aim of this study, only data from the pre-and post-intervention surveys were used . The data from the pre-intervention survey was used to assign participants from the experimental group to the belonging version of the intervention within the TIIM app, to access the demographics and to measure depression scores at baseline (T0). Moreover, depression and engagement scores were also taken from the post-intervention survey (T1).

The software IBM SPSS Statistics (Version 25) was used to analyze the data. A *p*-value below 0.05 was assumed to be significant. First, the individual identity code created by the participants was used to merge the data from the pre-and post-intervention surveys into a final dataset. The final data set was cleared by participants that had missing items in the pre-or post-survey intervention. Moreover, non-relevant data like the items from the GAD-7 and the MHC-SF questionnaire, as well as participants that did not start with the intervention within the TIIM app were excluded from the final sample. The data for this study was downloaded on the 27th of April 2022. At this point, it must be mentioned that data collection went beyond this date for the purpose of the larger research project. Descriptive statistics were used to analyze the participant's age, gender, nationality, and occupation. A t-test was used to determine the distribution of the different descriptives. Items belonging to the PHQ-9 (T0/T1) and the TWEETS questionnaire (T1) were summed up into a single scale score.

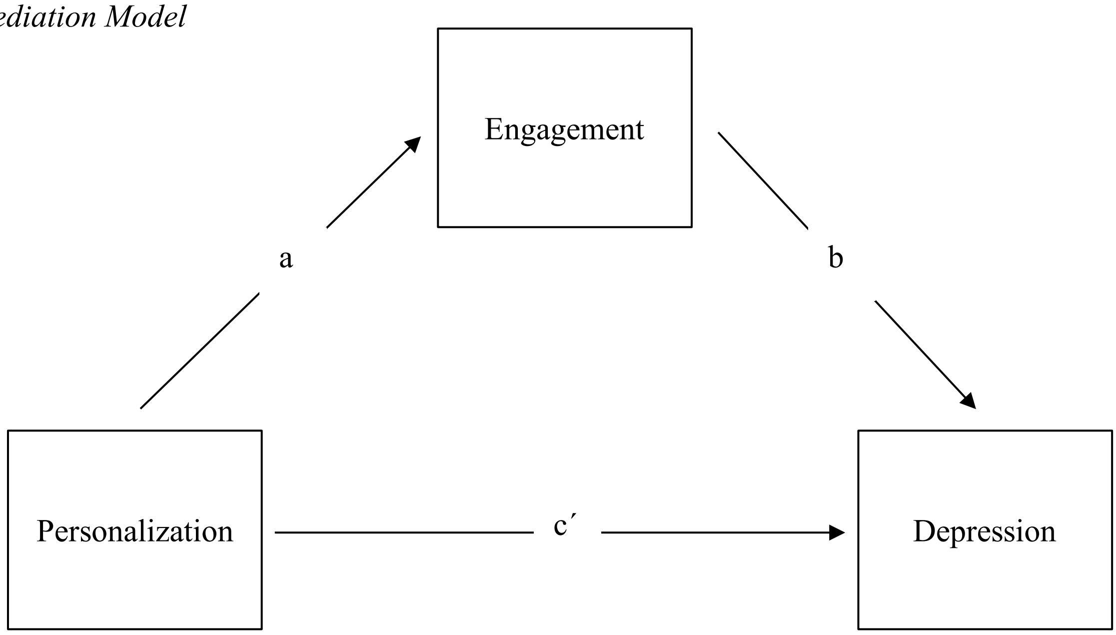
An independent sample t-test was used to investigate the differences between personalization and the effectiveness of the intervention at baseline and post-intervention. In addition, it gave insights into the relationship between personalization and engagement post-survey. For the analysis, the conditions of personalization (experimental/control group) were treated as categorical, independent variables. The depression and engagement scores were used as the dependent outcome variables.

A one-way repeated measures ANOVA was used to check whether there was a main effect of time on depression and an interaction effect of time*personalization on depression (T0-T1). The assumptions of normality and sphericity were checked and accepted. The Shapiro-Wilk tests showed non-normally distributed PHQ-9 scores for pre ($W(175) = .91$, $p = .00$) and post ($W(175) = .93$, $p = .00$) measurements. Nevertheless, the analysis was still conducted because the distribution was close to normal. Field (2018) and Pallant (2011) agrued that ANOVA is robust to this kind of violation. For the analyses, the conditions of personalization (experimental/control group were treated as categorical, independent variables.

Regarding the third research question, a mediation model (Figure 5) was used to investigate whether engagement mediates the relation between personalization as a condition and depression scores as the outcome variable. The IBM SPSS Statistics extension PROCESS version 3.5 by Hayes (2017) was used for this purpose. Bootstrap was set up to 5000 with a confidence interval of 95%. The bootstrap was used to generate a confidence interval around the indirect effects and to test them (Hesterberg, 2011). In addition, bootstrapping was used to reduce type one errors and has high statistical power for non-normal distributions (Cameron et al., 2008). Figure 5 shows the mediation model. Path *a* represents the effect of personalization on engagement. Path *b* displays the effect of engagement on the effectiveness of the intervention measured in depression scores. Path *c´* shows the direct effect of personalization on the effectiveness of the intervention after accounting for engagement.

Figure 5

Mediation Model



3. Results

3.1 Descriptive Statistics

Table 2 presents the sociodemographic data for the whole sample and for the two conditions control and experimental group. The total sample consisted of 176 participants. There were 87 participants in the control group, 88 participants in the experimental group and one missing value. The total sample had an age range between 18 and 31 with a mean age of 20.38 (*SD*= 1.84). Nearly three-quarters of the participants were female 72.6% and 27.4% identified as being male. Expect one missing value, all participants were students (99.4%). Regarding the distribution of the nationality, the majority of the participants were German (57.7%), 26.9% were Dutch, and 15.4% had other nationalities. The other nationalities varied from the US, South America (Brazil), Europe (Italy, Finland), and eastern countries like Russia, Indonesia, and Kazakhstan. In table 2, it can be observed that there are no big noticeable differences between the conditions control group and experimental group within the descriptives gender, age, occupation, and nationality. A t-test showed that the variance was equal within the categories gender (t(174) = -.26, p= .603), nationality (t(174) = -.28, p= .06), and occupation (all participants were students). The category age was not equally distributed (t(173) = -1.94, p= .002). However, most of the participants in both conditions were below the age of 26.

Table 2

Demographic information per condition and for the total sample

Characteristic		Control group		Experimental group		Total	
		n	%	n	%	n	%
Participant		87	49.4	88	50	176	100
	Missing					1	0.6
Gender							
	Male	25	28.7	23	26.1	48	27.4
	Female	62	71.3	65	73.9	127	72.6
	Other	-	-	-	-	-	-
	Missing					1	0.6
Age							
	<20	59	67.8	50	57.5	109	62.6
	21-25	28	32.2	35	40.2	63	36.2
	26-30	-	-	1	1.1	1	.6
	31-35	-	-	1	1.1	1	.6
	>35	-	-	-	-	-	-
	Missing					1	1.1
Occupation							
	Student	87	100	88	100	175	99.4
	Missing					1	0.6
Nationality							
	Dutch	26	29.9	21	23.9	47	26.9
	German	46	52.9	55	62.5	101	57.7
	Other	15	17.2	12	13.6	27	15.4
	Missing					1	0.6

3.2 Intervention effectiveness

Table 3 displays the mean scores and standard deviation of the variables depression and engagement for the control group as well as the experimental group. Overall, the outcomes of the independent t-test did not show any significant differences in the conditions.

Table 3

Descriptive statistics and independent t-test results for depression and engagement

	Condition						95% CI for Mean Difference		
	Control			Experimental					
	M	SD	n	M	SD	n		t	df
Depression T0	8.61	4.26	87	8.62	4.64	88	-1.3, 1.3	-.04	173
Depression T1	7.39	4.37	87	7.04	4.42	88	-.97, 1.6	.52	173
Engagement T1	31.63	6.41	87	31.15	6.52	88	-1.4, 2.4	.49	173

Note. M= means. SD= standard deviation. p>.05 for the control and experimental group for both outcome measures

3.2.1 Research question 1

With regard to research question one, – *Does the personalization of a 2-week Digital Health Intervention for depression and wellbeing directly influence the effectiveness of digital health interventions?* – the responses for depression covered almost the whole range of the PHQ-9 scale. The depression scores ranged from 1 to 24 (4-point Likert scale, 9 items). No significant difference in depression scores has been found between the conditions at baseline measurements (T0) t(173) = -.04, p= .97. The 88 participants from the experimental group (*M* = 8.62, *SD* = 4.64) compared to the 87 participants in the control group (*M* = 8.61, *SD* = 4.26) did not demonstrate significantly lower scores in depressive symptoms. The same results were found post measurements (T1) t(173) = .52, p= .60. In this case the experimental group (*M* = 7.04, *SD* = 4.42) compared to the control group (*M* = 7.39, *SD* = 4.37) did not demonstrate any significant differences either. However, looking at the results, it can be observed that the mean scores for depression are slightly lower at T1 than at T0.

The one-way repeated measures ANOVA that was carried out afterward – after the researcher investigated the results of the t-tests – showed a significant main effect of time on change in depression scores, Wilks lambda = .908 $F(1,173)$ =17.56, p < .001. However, no statistically significant interaction effect could be found for time and condition (experimental group vs. control group) on change in depression scores, Wilks lambda = .998 $F(1,173) = .35$, $p = .55$. This indicates that over time personalized interventions (experimental group) did not have a larger effect on depression scores than randomly assigned interventions (control group).

3.2.2 Research question 2

Regarding research question two, – *Does the personalization of a 2-week Digital Health Intervention for depression and wellbeing directly influence the participants' level of engagement?* – the 88 participants from the experimental group who received a personalized intervention (*M* = 31.15, *SD* = 6.52) did not show significantly higher engagement scores compared to the 87 participants in the control group who received a random intervention (*M* = 31.63, *SD* = 6.41). Therefore, no significant effect has been found in engagement scores between the conditions, t(173) = .495, p= .62. The mean scores of the engagement score measured by the TWEETS questionnaire scale ranged from 11 to 45 (5-point Likert scale, 9 items) within the whole sample.

3.3 Engagement as mediation

3.3.1 Research question 3

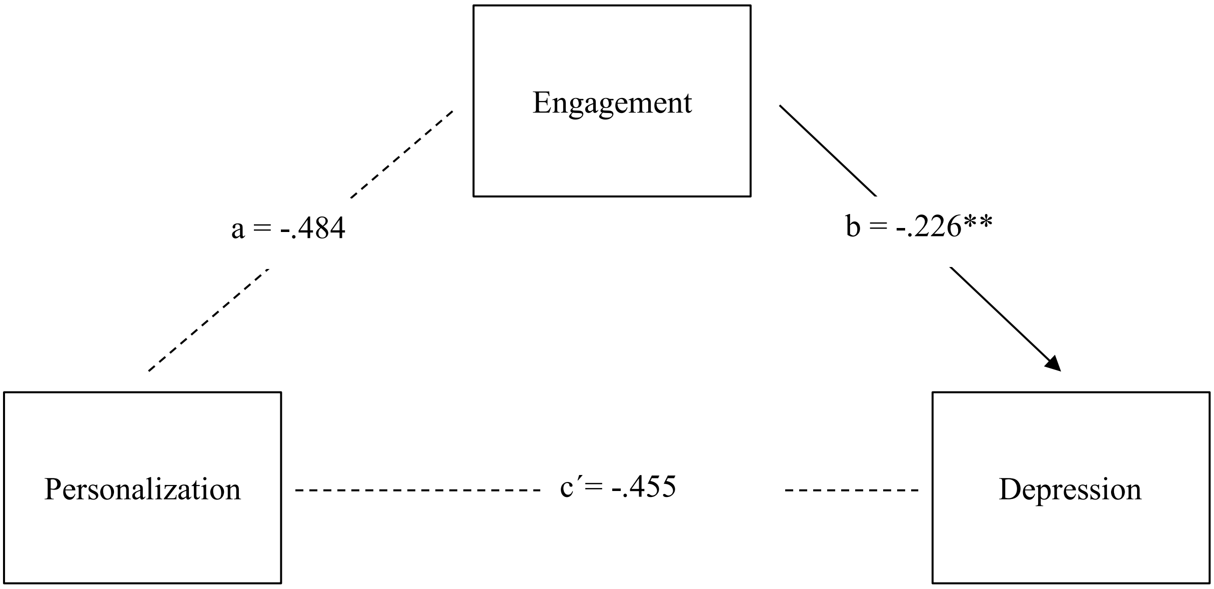
Looking at research question three, – *Does engagement mediate the relationship between personalization and the overall effectiveness of a 2-week Digital Health Intervention for depression and wellbeing?* – table 4 provides the results of the simple mediation analysis. In general, no mediation effect of engagement was found. The results show that there was no total effect (path *c*). Personalized digital health interventions did not predict a reduction in depression symptoms compared to the control group where participants received a random version of the DHI (*B* = -.34, *SE*= .66, *p*= .60). The results indicate that engagement did not significantly mediate the relationship between the two different conditions, control group and experimental group, and the depression outcome measures (Figure 6). A confidence interval of 95% based on 5000 bootstrap samples confirmed that zero was included within the indirect effect (control vs. experimental: [-.35, .55]). Looking at the *a* path, the results show that the control group and experimental group had no significant effect on engagement (*B* = -.48, *SE*= .98, *p*= .62). There were some significant results found in path *b*. DHI engagement was negatively related to depression scores meaning that higher engagement scores were related to lower depression scores and vice versa (*B* = -.23, *SE*= .04, *p*= <.01). Lastly, the results did not show any significant direct effects (path *c´*). More precisely the study conditions did not have an impact on depression outcome scores (*B* = -.45, *SE*= .63, *p*= .47).

Table 4
Simple mediation analysis of the effect control vs. experimental group, mediated by DHI engagement

Condition	Mediator	a	b	Relative total effect c	Relative direct effect c´	Relative indirect effect *a x b* (95% CI) [a]
Control vs Experimental	Engagement	-.48	-.23**	-.34	-.45	.11 (-.35, .55)

Note. [a] Bias corrected bootstrap results for the indirect effect *a x b* (5000 resamples). ***p* < .01.

Figure 6
Mediation model including the results of the mediation analysis



Note. Dotted lines represent non-significant results; bold lines represent significant results. ***p* < .01.

4. Discussion

4.1 Main findings

This study mainly aimed to investigate two different objectives. On the one hand, it was aimed to analyze to what extent the personalization of digital health interventions influences the participants engagement (RQ1) and the intervention effectiveness (RQ2). It was hypothesized that personalized versions of DHIs would lead to higher levels of engagement, and a stronger reduction of depressive symptoms compared to randomly assigned versions of DHIs. On the other hand, it was investigated if engagement mediates the relationship between personalized DHIs and the effectiveness of DHIs accessed through the reduction of depressive symptoms.

The findings of this study did not show differences on average for the two conditions, namely personalized and non-personalized DHI. Results did show an improvement over time in depression scores indicating that digital health interventions in the form of mobile apps might help to improve mental well-being within a short period of time. The small to medium effect size found for the main effect of time on effectiveness can be accepted since this study included an unguided and short intervention.

Similar to the t-test an interaction effect between time and condition (personalized vs. non-personalized DHI) on depression scores was not found. The results of the one-way repeated measures ANOVA and the t-test were used in section 4.2 in order to discuss potential reasons for the non-significant results.

4.2 Research questions 1 and 2

4.2.1 Personalization of the intervention

Research questions one and two aimed to determine if the personalization of DHIs could be used to make these more engaging and more effective for the users. The data analysis showed that personalization did neither significantly influence the engagement level of DHI users nor the effectiveness of the DHI.

In the current study, participants from the experimental group received a personalized intervention. The personalization procedure was used to assign the participants to the best-fitting version of the DHI which might have not worked well enough. During the data analysis, it was discovered that nearly two-thirds of the total dataset – 164 out of 254 participants derived from the total data set at the 6th of June 2022 – had the same scores between two or three options for one factor. For example, participants had the same score for the *feedback* option plain written text and words spoken by a human counselor in a pre-recorded video. In this case, an online tool assigned the participants to one of the two options randomly. This had the consequence that the intervention was not completely personalized towards the needs and preferences of the participant. On the one hand, it can be hypothesized, that participants who had the same scores in two or three options for one factor either liked both options equally and could not decide which one they prefer. On the other hand, the different options of the three factors *content, feedback,* and *design* might not have been explained well enough which could have resulted in participants not filling out the questionnaire consciously enough. Data analysis showed that five participants from the non-personalized group received a random version of the

DHI that fitted their personal needs and preferences. Due to this small number, it cannot be hypothesized that these findings was the reason for the non-significant results found in research questions one and two. To the knowledge of the researcher, recent studies have not been confronted with these issues (Dennison et al., 2014; Tang et al., 2015).

Past research that determined the effectiveness of treatment for major depression could confirm that user preference was related to the engagement level of the participants (Gelhorn et al., 2011; Kwan et al., 2010). However, a study from Brenes et al. (2021) could not find a significant effect for preference on engagement in cognitive behavioral therapy and yoga interventions among older adults. For future research, it is recommended to include the factor preference in order to further investigate if the conscious decision about a preferred version of an intervention might increase the level of engagement within the participants. For example, at the beginning of the intervention participants could receive an explanation for every option (competitive gamification, non-competitive gamification, and no gamification elements) of the factor *design*. After that, participants with a personalized DHI would be able to choose one option for one factor by themselves. This could prevent the discussed issue ofhaving equal scores for different options within one factor because the participants would need to actively decide on one option themselves. During the current study this whole process was more unconscious for the participants because the person did not know that the personalization procedure was used to determine the best fitting intervention. It can be hypothesized, that a stronger integration of the participant within the personalization procedure might have a different effect on the level of engagement.

4.2.2 Factors of the Intervention (TIIM App)

Another point that needs to be discussed is the presentation of the factors *content, feedback,* and *design*. Kelders et al. (2019) and Perski et al. (2016) hypothesized that tailored DHIs increased the motivation to use the app. This can be reached by designing the content and feedback of the DHI in a way that it meets the values and beliefs of the user (Hayland & Whalley, 2008; Hayland & Whalley, 2009; Kelders, 2019; Zagorscak et al., 2020). As mentioned before, the current study was part of a larger pilot study. Certain aspects like the intervention within the TIIM app were not fully professionalized. More precisely, the intervention and technological factors of the current study namely *content, feedback,* and *design* were static, simple, and impersonal. For example, the feedback messages were the same for all participants (impersonal). Moreover, the competitive gamification design was illustrated by bricks moving from left to right with the ongoing progress of the participant (simple). This gave limited options for a tailored intervention that fitted the participants' needs and beliefs.

Within the factor *feedback,* participants continuously received generic and extrinsic feedback which was not based on their individual performance. As mentioned earlier, the daily feedback messages provided by the DHI were always the same between the different feedback options and for all participants. The OPTIMAL theory (Optimizing Performance through Intrinsic Motivation and Attention for Learning) states that success in performance and engagement to learn new things can be increased by autonomy and intrinsic motivation rather than extrinsic motivation (Wulf & Lewthwaite, 2016). An older study by Curry et al. from 1991 investigated the effects of intrinsic and extrinsic motivation interventions with a self-help smoking cessation program. Curry et al. (1991) promoted intrinsic motivation through personalized feedback and external motivation through financial incentives in the form of gift cards. It was found that personalized feedback increased intrinsic motivation and showed better outcomes as well. Consequently, feedback within DHI can increase intrinsic motivation as well. Therefore, it can be hypothesized that the feedback messages in the current study were not personalized nor tailored enough.

Looking at the factor *design* it can be said that the different design options of the intervention were rather simple than complex. Sieverink et al. (2017) hypothesized that the active participation in DHIs is dependent on the usefulness of the application elements that help to successfully reach their own needs and goals. Within the current study the competitive gamification design was characterized by bricks that move from left to right with the ongoing progress. Here, it can be questioned if this is useful and induces competitiveness within the user. The non-competitive gamification design showed a cyclist on a bicycle tour. It can be hypothesized, that especially these two designs were not different enough. Research has found that learning material that consists of gamification and storyline elements (non-competitive gamification) can be very engaging and motivating. It was found that competitive gamification elements in learning materials foster deep and frequent processing (Mayer & Mayer, 2005). Moreover, the term content gamification describes mechanisms in which the learning material will be presented in a storyline that aims to make the learning material more engaging and interesting for the learner (Mayer & Mayer, 2005). In the current study, the content of the DHI can be considered as learning material since the users learn new techniques on how they can better handle their depressive symptoms. Therefore, it can be argued that the presentation of the content in form of the design was insufficient in the context of being competitive and providing a storyline during the progress of the intervention.

Future research should design the intervention feedback in a more personal and individual manner. Feedback should be provided based on the input of the participant. The feedback could entail praise for a specific goal the participant has set and might explain why this is a good goal (Curry et al., 1991; Mumm & Mutlu, 2011). In addition, a feature could be added in which the participant has the opportunity to give feedback on their own performance.

Furthermore, future research should create a more detailed design of the DHI. The difference between the design of competitive gamification and non-competitive gamification should be larger. Improved versions of DHIs should include elements that are either more competitive or more storytelling than the ones of the current versions. Increasing the difference between the different intervention designs might have the effect of being more effective and being more engaging because the participants might experience the aspect of competition or storyline more intensely. In this case, it also needs to be mentioned that professional intervention designs that include more gamification, competitiveness, and storyline are also more expensive and time-consuming. In addition, future studies could include a feature in which the participants would have the option to adjust their intervention after a certain period. For example, changing from a competitive gamification design to a noncompetitive gamification design after one week of participation. Adjusting the intervention after a certain period might improve the personalization of the intervention, increase the perceived autonomy of the participants, and can lead to a higher level of engagement (Kelders et al., 2020a; Wulf & Lewthwaite, 2016).

4.3 Research question three

4.3.1 Mediation Model

Research question three investigated whether engagement mediates the relationship between personalization and the overall effectiveness of DHIs. It can be argued that the personalization procedure did not work well. Therefore, no differences in engagement, outcomes, and no mediating effect was found within the data analysis. The only significant results found were that engagement was related to depression scores. Currently, there are no other studies with a comparable mediation model. The current study used the conceptual mediation model from the grant application form (research proposal by Kelders, 2017). It can be hypothesized, that the investigation of the personalization of the factors *design*, *feedback*, and *content* as individual independent variables could have shown more detailed insights into which factors are especially important to increase engagement. Furthermore, it might be the case that other factors have a bigger impact on engagement than personalization. Borghouts et al. (2021) did a literature review in order to conceptualize the barriers and facilitators that influence user engagement in DHIs. In total, they could identify 16 factors that had a positive effect on user engagement. Factors like level of guidance, type of content, or personal traits of the users were identified as facilitators. Moreover, a study from 2016 found that patient's engagement was a mediator for the relationship between the perceived ability of healthcare professionals in motivating type 2 diabetes patients and activation of self-management (Graffigna et al., 2016).

Future research should further improve the personalization approach in order to prevent failures as in the current study. In addition, they should concentrate on other facilitators of engagement. It is recommended to take existing significant facilitators as guidelines for future study designs (Borghouts et al., 2021; Woldaregay et al., 2018). A range of potential causing factors could give new important and interesting insights into the field of DHIs.

4.3.2 Negative correlation path b

Path *b* of the mediation model was the only result found to be significant within the mediation model. The significantly negative results indicate that participants with high engagement scores had low depression scores and vice versa. This means that participants with a higher level of depressive symptoms were less engaged with the DHI than participants with a low level of depressive symptoms. These results can be interpreted in two different ways. On the one hand, it can be argued that participants with a lower depression score on post-intervention are the ones for whom the intervention was effective. Therefore, it can be hypothesized that participants that were highly engaged with the app also experienced positive effects after 2 weeks of using the intervention. On the other hand, it can be hypothesized, that the intervention was not effective and that participants with low depression scores at post-intervention already had these low depression scores at baseline. In addition, participants with a high level of depressive symptoms might have had high expectations for the DHI to minimize their depressive symptoms. Not meeting these expectations could have resulted inthese participants not engaging with the app. Previous research shows that expectation and satisfaction have an impact on the adherence on DHIs (Beattie et al., 2009; Boß et al., 2016). In contrast, nearly all participants were students from the University of Twente. The sample was very specific because the majority of the study were young female students from Germany. Students with a background in behavioural or social science that participated in the study might have had low level of depression scores but high engagement scores because they were interested in the topic of digital health interventions and not in receiving sufficient treatment.

Another reason could be that the participants did not receive any introduction, support, or guidance from a health professional. The different exercises required a certain amount of self-initiative and self-management. It may be that a certain amount of the participants didn't have any prior experience with therapy or the treatment of mental health problems like depressive symptoms. Therefore, it can be hypothesized that some participants with high level of depressive symptoms were not able to process the experiences during the DHI without the support of a health professional and were therefore not engaged with the app.

Future research that aims to further investigate a specific population, in this case people with depressive symptoms, should improve the study design by addressing the specific needs of this population. Pilot tests and pilot surveys could be used to identify the expectations that are especially important for this target population (Secomb & Smith, 2011). In addition, participants should receive additional support or at least have the opportunity to contact a health care professional. Concentrating on one specific target population could become beneficial for the aspect of practical implications. Taking the target population of young female students as an example it can be argued that in case of significant findings the app could be recommended by a student psychologist, study counselor, or the University homepage as soon as the target population experiences and reports symptoms of stress, anxiety, or depression (Stewing, 2021).

4.4 Strengths and limitations

This study includes two characteristics that can be interpreted as a strength and limitation at the same time. The demographic characteristics of the sample were very specific. All participants were students and most of them were female students from Germany below the age of 20. It can be considered as a limitation because the finding of the current study cannot be applied to the overall society which entails a broader variety of characteristics. Nevertheless, the results found in this study can be applied to a particular population within the society which can be considered a strength. The findings give the opportunity to gain specific insights into a population that was not studied before in this particular context. The second aspect is the study design of the current study. Based on the critique and the implications for future research from Stewing (2021) it can be argued that it is a strength that the current study included a control group. The participants were randomly assigned to one of the two conditions namely the personalized or non-personalized group. This made the results found in this study statistically more solid. A limitation was that many participants within the personalized group received a random version of the intervention as well. Participants that had equal scores in the variations of one factor received an intervention version randomly. Beforehand, it was not considered that a high number of randomly assigned intervention versions within the personalized group could influence the degree of accuracy. Therefore, it can be argued that the personalization procedure as a whole was not sufficient enough.

This study had several strengths. One of them was the structure of the overall study including the data collection as well as the intervention within the TIIM app itself. The study was conducted completely remotely. Considering that the topic was about digital health interventions which should aim to reduce cost, geographical-, and time barriers it made sense to create a study that is completely digital. This made the user experience for the participants more realistic and had a positive impact on the expressive power of the results. Another strong point was the variety of intervention versions. The intervention "Learn to flourish" within the TIIM app contained some technical issues. Nevertheless, under consideration that this intervention was a pilot project, it can be argued that it gave the opportunity for personal adjustments for the user to improve their own mental well-being. The 27 different intervention versions and the overall design of the app made the experience for the participant more realistic. Data was collected at three different time points with four different questionnaires. It must be mentioned that not all data from the different measurement points and questionnaires were used in this study. However, it is a strength to have this value of the collected data since the head of the larger research project will use this data.

Another limitation of this study was the technical problems experienced by the participants during the intervention. The researcher could identify 41 cases reported via email in which the users either experienced technical issues with the app or had problems with understanding the overall procedure of the study. In addition, complications with the SONA system had the consequence that participants reported being irritated about the structure of the study. All these issues could have had a negative effect on the level of engagement and the overall satisfaction with the app.

5. Conclusion

This study gave some interesting insights into the role of personalization and engagement in a digital health intervention for people with mental issues. No significant results were found for the influence of personalization of DHIs on the engagement level as well as the level of depressive symptoms. However, it was found that depression scores significantly improve over time. A mediating role of engagement on the relationship between personalization and the outcome measure depressive symptom scores was not found. It was found that high engagement scores are related to reduced depressive symptoms at post measurements. Possible reasons for these results could be issues with the personalization procedure as well as technical problems during the data collection. Future research is advised to improve the personalization of DHIs for example by increasing the user's autonomy. Besides that, adding further potential causing factors for engagement could lead to other important findings for the establishment of solid eHealth technologies.

6. References

Andersson, G., & Cuijpers, P. (2008). Pros and cons of online cognitive–behavioural therapy. *The British Journal of Psychiatry*, *193*(4), 270-271. doi:10.1192/bjp.bp.108.054080

Andrade, L. H., Alonso, J., Mneimneh, Z., Wells, J. E., Al-Hamzawi, A., Borges, G., ... & Kessler, R. C. (2014). Barriers to mental health treatment: results from the WHO World Mental Health surveys. *Psychological medicine*, *44*(6), 1303-1317. DOI:10.1017/S0033291713001943

Andrews, G., Cuijpers, P., Craske, M. G., McEvoy, P., & Titov, N. (2010). Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. *PloS one*, *5*(10), e13196 https://doi.org/10.1371/journal.pone.0013196

Aref-Adib, G., McCloud, T., Ross, J., O'Hanlon, P., Appleton, V., Rowe, S., ... & Lobban, F. (2019). Factors affecting implementation of digital health interventions for people with psychosis or bipolar disorder, and their family and friends: a systematic review. *The Lancet Psychiatry*, *6*(3), 257-266. https://doi.org/10.1016/S2215-0366(18)30302-X

Beattie, A., Shaw, A., Kaur, S., & Kessler, D. (2009). Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. *Health expectations*, *12*(1), 45-59. DOI: 10.1111/j.1369-7625.2008.00531.x

Boß, L., Lehr, D., Reis, D., Vis, C., Riper, H., Berking, M., & Ebert, D. D. (2016). Reliability and validity of assessing user satisfaction with web-based health interventions. *Journal of medical Internet research*, *18*(8), e5952. DOI: 10.2196/jmir.5952

Borghouts, J., Eikey, E., Mark, G., De Leon, C., Schueller, S. M., Schneider, M., ... & Sorkin, D. H. (2021). Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. *Journal of medical Internet research*, *23*(3), e24387. DOI: 10.2196/24387

Brenes, G. A., Clary, H. M. M., Miller, M. E., Divers, J., Anderson, A., Hargis, G., & Danhauer, S. C. (2021). Predictors of preference for cognitive-behavioral therapy (CBT) and yoga interventions among older adults. *Journal of Psychiatric Research*, *138*, 311-318. https://doi.org/10.1016/j.jpsychires.2021.03.055

Bundespsychotherapeutenkammer (2018) Ein Jahr nach der Reform der Psychotherapie-Richtlinie: Wartezeiten 2018. https://www.bptk.de/publikationen/bptk-studie/.

Burton, C. M., & King, L. A. (2004). The health benefits of writing about intensely positive experiences. *Journal of research in personality*, *38*(2), 150-163. https://doi.org/10.1016/S0092-6566(03)00058-8

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The review of economics and statistics*, *90*(3), 414-427.

Cole-Lewis, H., Ezeanochie, N., & Turgiss, J. (2019). Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. *JMIR formative research*, *3*(4), e14052. doi: 10.2196/14052

Curry, S. J., Wagner, E. H., & Grothaus, L. C. (1991). Evaluation of intrinsic and extrinsic motivation interventions with a self-help smoking cessation program. *Journal of consulting and clinical psychology*, *59*(2), 318.

Dennison, L., Morrison, L., Lloyd, S., Phillips, D., Stuart, B., Williams, S., ... & Yardley, L. (2014). Does brief telephone support improve engagement with a web-based weight management intervention? Randomized controlled trial. *Journal of medical Internet research*, *16*(3), e95. DOI: 10.2196/jmir.3199

Du, S., Liu, W., Cai, S., Hu, Y., & Dong, J. (2020). The efficacy of e-health in the self-management of chronic low back pain: A meta analysis. *International journal of nursing studies*, *106*, 103507. https://doi.org/10.1016/j.ijnurstu.2019.103507

Erbe, D., Eichert, H. C., Riper, H., & Ebert, D. D. (2017). Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. *Journal of medical Internet research*, *19*(9), e306. doi: 10.2196/jmir.6588

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (Fifth ed.). SAGE Publications Ltd.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry*, *16*(3), 287-298.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*, *4*(2), e7785. doi: 10.2196/mental.7785

Fortuna, K. L., Torous, J., Depp, C. A., Jimenez, D. E., Areán, P. A., Walker, R., & Bartels, S. J. (2019). A future research agenda for digital geriatric mental healthcare. *The American Journal of Geriatric Psychiatry*, *27*(11), 1277-1285. https://doi.org/10.1016/j.jagp.2019.05.013

Gelhorn, H. L., Sexton, C. C., & Classi, P. M. (2011). Patient preferences for treatment of major depressive disorder and the impact on health outcomes: a systematic review. *The primary care companion for CNS disorders*, *13*(5), 27083. DOI:10.4088/PCC.11r01161

Glied, S., McCormack, S., & Neufeld, A. (2003). Women with depression: Financial barriers to access. *Professional Psychology: Research and Practice*, *34*(1), 20. DOI: 10.1037/0735-7028.34.1.20

Gorst, S. L., Armitage, C. J., Brownsell, S., & Hawley, M. S. (2014). Home telehealth uptake and continued use among heart failure and chronic obstructive pulmonary disease patients: a systematic review. *Annals of Behavioral Medicine*, *48*(3), 323-336. https://doi.org/10.1007/s12160-014-9607-x

Graffigna, G., Barello, S., Bonanomi, A., & Menichetti, J. (2016). The motivating function of healthcare professional in eHealth and mHealth interventions for type 2 diabetes patients and the mediating role of patient engagement. *Journal of diabetes research*, *2016*. https://doi.org/10.1155/2016/2974521

Groeneveld, B. S. (2020). *Talking the walk: Applying data-driven patient profiles in the Design of tailored services in orthopaedics* (Doctoral dissertation, Delft University of Technology). DOI 10.4233/uuid:7c866494-52b6-4a53-baff-f6b24443ef14

Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?--a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences* (pp. 3025-3034). Ieee.

Handley, T. E., Kay-Lambkin, F. J., Inder, K. J., Attia, J. R., Lewin, T. J., & Kelly, B. J. (2014). Feasibility of internet-delivered mental health treatments for rural populations. *Social psychiatry and psychiatric epidemiology*, *49*(2), 275-282 https://doi.org/10.1007/s00127-013-0708-9

Hayes, A. F. (2017). Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition: A Regression-Based Approach *(Methodology in the Social Sciences)* (Second ed.). The Guilford Press.

Hedman, E., Ljótsson, B., & Lindefors, N. (2012). Cognitive behavior therapy via the

Internet: a systematic review of applications, clinical efficacy and cost–effectiveness. *Expert review of pharmacoeconomics & outcomes research*, *12*(6), 745-764. https://doi.org/10.1586/erp.12.67

Hesterberg, T. (2011). Bootstrap. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(6), 497-526. https://doi.org/10.1002/wics.182

Hollis, C., Morriss, R., Martin, J., Amani, S., Cotton, R., Denis, M., & Lewis, S. (2015). Technological innovations in mental healthcare: harnessing the digital revolution. *The British Journal of Psychiatry*, *206*(4), 263 265. https://doi.org/10.1192/bjp.bp.113.142612

Hyland, M. E., & Whalley, B. (2008). Motivational concordance: An important mechanism in self-help therapeutic rituals involving inert (placebo) substances. *Journal of psychosomatic research*, *65*(5), 405-413. https://doi.org/10.1016/j.jpsychores.2008.02.006

Jacobi, F., Höfler, M., Strehle, J., Mack, S., Gerschler, A., Scholl, L., & Wittchen, H. U. (2014). Psychische Störungen in der Allgemeinbevölkerung. *Der Nervenarzt*, *85*(1), 77-87. DOI 10.1007/s00115-013-3961-y

Kelders, S. M. (2019). Design for engagement of online positive psychology interventions. In *Positive psychological intervention design and protocols for multi-cultural contexts* (pp. 297-313). Springer, Cham. DOI: 10.1007/978-3-030-20020-6_13

Kelders, S. M., & Kip, H. (2019). Development and initial validation of a scale to measure engagement with eHealth technologies. In *CHI EA 2019 - Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* [3312917] Association for Computing Machinery (ACM). https://doi.org/10.1145/3290607.3312917

Kelders, S. M., Kip, H., & Greeff, J. (2020b). Psychometric evaluation of the Twente engagement with Ehealth Technologies Scale (TWEETS): evaluation study. *Journal of Medical Internet Research*, *22*(10), e17757. DOI: 10.2196/17757

Kelders, S. M., Van Zyl, L. E., & Ludden, G. D. (2020a). The concept and components of engagement in different domains applied to ehealth: a systematic scoping review. *Frontiers in psychology*, *11*, 926. https://doi.org/10.3389/fpsyg.2020.00926

Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual review of public health*, *34*, 119-138. doi: 10.1146/annurev publhealth-031912-114409

Keyes CLM. Atlanta: Brief Description of the Mental Health Continuum Short Form (MHC-SF).2014. https://www.aacu.org/sites/default/files/MHC-SFEnglish.pdf. Accessed 08.12.15.

Klein, J. P., Berger, T., Schröder, J., Späth, C., Meyer, B., Caspar, F., ... & Hohagen, F. (2016). Effects of a psychological internet intervention in the treatment of mild to moderate depressive symptoms: results of the EVIDENT study, a randomized controlled trial. *Psychotherapy and psychosomatics*, *85*(4), 218-228. https://doi.org/10.1159/000445355

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, *16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kwan, B. M., Dimidjian, S., & Rizvi, S. L. (2010). Treatment preference, engagement, and clinical improvement in pharmacotherapy versus psychotherapy for depression. *Behaviour research and therapy*, *48*(8), 799-804. https://doi.org/10.1016/j.brat.2010.04.003

Lal, S. (2019). E-mental health: Promising advancements in policy, research, and practice. *Healthcare Management Forum*, *32*(2), 56–62. https://doi.org/10.1177/0840470418818583

Lukas, C. A., Eskofier, B., & Berking, M. (2021). A gamified smartphone-based intervention for depression: Randomized controlled pilot trial. *JMIR Mental Health*, *8*(7), e16643. DOI: 10.2196/16643

Mack, S., Jacobi, F., Gerschler, A., Strehle, J., Höfler, M., Busch, M. A., & Wittchen, H. U. (2014). Self-reported utilization of mental health services in the adult German population–evidence for unmet needs? Results of the DEGS1-Mental Health Module (DEGS1-MH). *International Journal of Methods in Psychiatric Research*, *23*(3), 289-303. https://doi.org/10.1002/mpr.1438

Malla, A., Joober, R., & Garcia, A. (2015). “Mental illness is like any other medical illness”: a critical examination of the statement and its impact on patient care and society. *Journal of psychiatry & neuroscience: JPN*, *40*(3), 147. DOI: 10.1503/jpn.150099

Mangiapane, S., Zhu, L., Czihal, T., & von Stillfried, D. (2020). Veränderung der vertragsärztlichen Leistungsinanspruchnahme während der COVID-Krise. *Zentralinstitut für die kassenärztliche Versorgung in der Bundesrepublik Deutschland*.

Mayer, R., & Mayer, R. E. (Eds.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge university press.

Meadows, G. N., Enticott, J. C., Inder, B., Russell, G. M., & Gurr, R. (2015). Better access to mental health care and the failure of the Medicare principle of universality. *Medical Journal of Australia*, *202*(4), 190-194. https://doi.org/10.5694/mja14.00330

Mennis, J., Stahler, G. J., & Baron, D. A. (2012). Geographic barriers to community-based psychiatric treatment for drug-dependent patients. *Annals of the Association of American Geographers*, *102*(5), 1093-1103. https://doi.org/10.1080/00045608.2012.657142

Meyer, B., Bierbrodt, J., Schröder, J., Berger, T., Beevers, C. G., Weiss, M., ... & Klein, J. P. (2015). Effects of an Internet intervention (Deprexis) on severe depression symptoms: Randomized controlled trial. *Internet Interventions*, *2*(1), 48-59. https://doi.org/10.1016/j.invent.2014.12.003

Molloy, A., & Anderson, P. L. (2021). Engagement with mobile health interventions for depression: A systematic review. *Internet Interventions*, *26*, 100454. https://doi.org/10.1016/j.invent.2021.100454

Mumm, J., & Mutlu, B. (2011). Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Computers in Human Behavior*, *27*(5), 1643-1650. https://doi.org/10.1016/j.chb.2011.02.002

National Institute for Public Health and the Environment. (June 4, 2020). Changes in mental health due to the coronavirus in the Netherlands in 2020, by category. In *Statista*. Retrieved December 02, 2021, from https://www.statista.com/statistics/1115803/dutch-mental-health-coronavirus/

O'connor, S., Hanlon, P., O'donnell, C. A., Garcia, S., Glanville, J., & Mair, F. S. (2016). Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. *BMC medical informatics and decision making*, *16*(1), 1-15. https://doi.org/10.1186/s12911-016-0359-3

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.

Pennant, M. E., Loucas, C. E., Whittington, C., Creswell, C., Fonagy, P., Fuggle, P., ... & Group, E. A. (2015). Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. *Behaviour research and therapy*, *67*, 1-18. https://doi.org/10.1016/j.brat.2015.01.009

Perski, O., Blandford, A., West, R., & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Translational behavioral medicine*, *7*(2), 254-267. https://doi.org/10.1007/s13142-016-0453-1

Radez, J., Reardon, T., Creswell, C., Lawrence, P. J., Evdoka-Burton, G., & Waite, P. (2021). Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. *European child & adolescent psychiatry*, *30*(2), 183-211. https://doi.org/10.1007/s00787-019-01469-4

Richardson, T., Stallard, P., & Velleman, S. (2010). Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. *Clinical child and family psychology review*, *13*(3), 275-290. https://doi.org/10.1007/s10567-010-0069-9

Sanders, C., Rogers, A., Bowen, R., Bower, P., Hirani, S., Cartwright, M., ... & Newman, S. P. (2012). Exploring barriers to participation and adoption of telehealth and telecare within the Whole System Demonstrator trial: a qualitative study. *BMC health services research*, *12*(1), 1-12. https://doi.org/10.1186/1472-6963-12-220

Saperstein, S. L., Atkinson, N. L., & Gold, R. S. (2007). The impact of Internet use for weight loss. *Obesity reviews*, *8*(5), 459-465. https://doi.org/10.1111/j.1467789X.2007.00374.x

Saxena, S., Thornicroft, G., Knapp, M., & Whiteford, H. (2007). Resources for mental health: scarcity, inequity, and inefficiency. *The lancet*, *370*(9590), 878-889. https://doi.org/10.1016/S0140-6736(07)61239-2

Sebri, V., & Savioni, L. (2020). An introduction to personalized eHealth. In *P5 eHealth: An Agenda for the Health Technologies of the Future* (pp. 53-70). Springer, Cham. https://doi.org/10.1007/978-3-030-27994-3_4

Secomb, J. M., & Smith, C. (2011). A mixed method pilot study: The researchers' experiences. *Contemporary nurse*, *39*(1), 31-35. https://doi.org/10.5172/conu.2011.39.1.31

Short, C. E., DeSmet, A., Woods, C., Williams, S. L., Maher, C., Middelweerd, A., ... & Crutzen, R. (2018). Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. *Journal of medical Internet research*, *20*(11), e9397. DOI: 10.2196/jmir.9397

Sigg, S., Lagerspetz, E., Peltonen, E., Nurmi, P., & Tarkoma, S. (2016). Sovereignty of the apps: there's more to relevance than downloads. *arXiv preprint arXiv:1611.10161*.

Sieverink, F., Kelders, S. M., & van Gemert-Pijnen, J. E. (2017). Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. *Journal of Medical Internet Research*, *19*(12), e402. https://doi.org/10.2196/jmir.8578

Stein, D. J., Phillips, K. A., Bolton, D., Fulford, K. W. M., Sadler, J. Z., & Kendler, K. S. (2010). What is a mental/psychiatric disorder? From DSM-IV to DSM-V. *Psychological medicine*, *40*(11), 1759-1765. https://doi.org/10.1017/S0033291709992261

Stewing, S. (2021). *The Impact and Role of Feedback and Engagement in a Digital Health Intervention for Depression* (Master's thesis, University of Twente).

Talbot, F. (2012). Client contact in self-help therapy for anxiety and depression: Necessary but can take a variety of forms beside therapist contact. *Behaviour Change*, *29*(2), 63-76 https://doi.org/10.1017/bec.2012.4

Tang, J., Abraham, C., Stamp, E., & Greaves, C. (2015). How can weight-loss app designers' best engage and support users? A qualitative investigation. *British journal of health psychology*, *20*(1), 151-171. https://doi.org/10.1111/bjhp.12114

Taylor-Rodgers, E., & Batterham, P. J. (2014). Evaluation of an online psychoeducation intervention to promote mental health help seeking attitudes and intentions among young adults: randomised controlled trial. *Journal of affective disorders*, *168*, 65-71. https://doi.org/10.1016/j.jad.2014.06.047

Topooco, N., Berg, M., Johansson, S., Liljethörn, L., Radvogin, E., Vlaescu, G., ... & Andersson, G. (2018). Chat-and internet-based cognitive–behavioural therapy in treatment of adolescent depression: randomised controlled trial. *BJPsych open*, *4*(4), 199-207. https://doi.org/10.1192/bjo.2018.18

Tyerman, J., Patovirta, A. L., & Celestini, A. (2021). How stigma and discrimination influences nursing care of persons diagnosed with mental illness: a systematic review. *Issues in mental health nursing*, *42*(2), 153-163. https://doi.org/10.1080/01612840.2020.1789788

van Agteren, J., Bartholomaeus, J., Steains, E., Lo, L., & Gerace, A. (2021). Using a Technology-Based Meaning and Purpose Intervention to Improve Well-being: A Randomised Controlled Study. *Journal of Happiness Studies*, 1-21. https://doi.org/10.1007/s10902-021-00383-4

Watson, S., Woodside, J. V., Ware, L. J., Hunter, S. J., McGrath, A., Cardwell, C. R., ... & McKinley, M. C. (2015). Effect of a web-based behavior change program on weight loss and cardiovascular risk factors in overweight and obese adults at high risk of developing cardiovascular disease: randomized controlled trial. *Journal of medical Internet research*, *17*(7), e177.

Wehrmeyer, S. (2021). *The Mediating Role of Digital Health Intervention (DHI) Engagement in the Relationship between Feedback Variations and Mental Well-Being* (Master's thesis, University of Twente).

Whalley, B., & Hyland, M. E. (2009). One size does not fit all: Motivational predictors of contextual benefits of therapy. *Psychology and Psychotherapy: Theory, Research and Practice*, *82*(3), 291-303. https://doi.org/10.1348/147608309X413275

Williams, N. (2014). The GAD-7 questionnaire. *Occupational medicine*, *64*(3), 224-224. https://doi.org/10.1093/occmed/kqt161

Wittchen, H. U., & Jacobi, F. (2005). Size and burden of mental disorders in Europe—a critical review and appraisal of 27 studies. *European neuropsychopharmacology*, *15*(4), 357-376. https://doi.org/10.1016/j.euroneuro.2005.04.012

Woldaregay, A. Z., Issom, D. Z., Henriksen, A., Marttila, H., Mikalsen, M., Pfuhl, G., ... & Hartvigsen, G. (2018, June). Motivational Factors for User Engagement with mHealth Apps. In *pHealth* (pp. 151-157).

Wulf, G., & Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic bulletin & review*, *23*(5), 1382-1414. https://doi.org/10.3758/s13423-015-0999-9

Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., ... & Blandford, A. (2016). Understanding and promoting effective engagement with digital behavior change interventions. *American journal of preventive medicine*, *51*(5), 833-842. https://doi.org/10.1016/j.amepre.2016.06.015

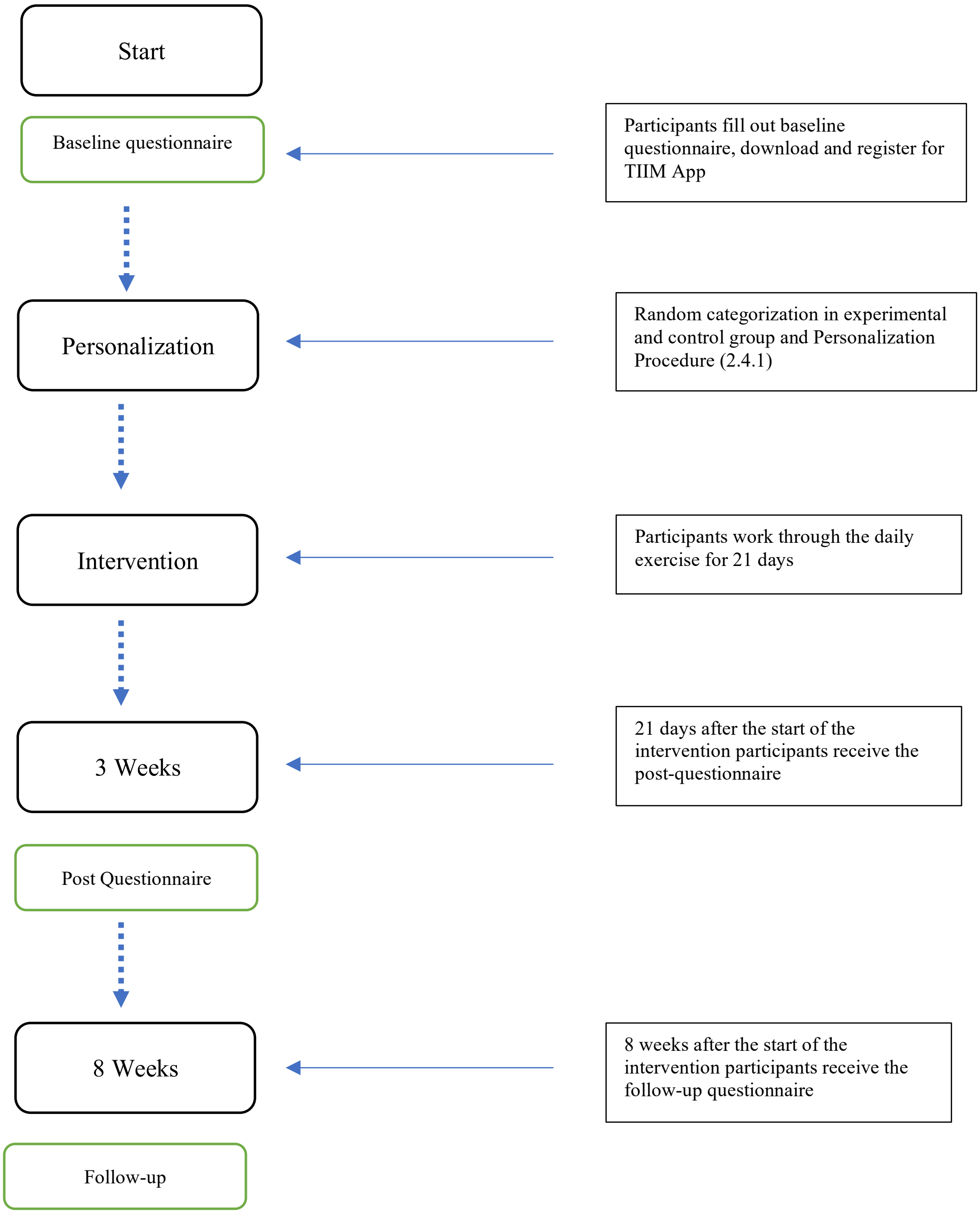
Zagorscak, P., Heinrich, M., Schulze, J., Böttcher, J., & Knaevelsrud, C. (2020). Factors Contributing to Symptom Change in Standardized and Individualized Internet- based Interventions for Depression: A Randomized-controlled Trial. *Psychotherapy*, *57*(2),237–251. https://doi.org/10.1037/pst0000276

Zhu, J., Dallal, D. H., Gray, R. C., Villareale, J., Ontañón, S., Forman, E. M., & Arigo, D. (2021). Personalization paradox in behavior change apps: Lessons from a social comparison-based personalized app for physical activity. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1-21. https://doi.org/10.1145/3449190

7. Appendices

Appendix A

Flow chart of intervention study including the measurement points baseline and post questionnaire.



Note. All questionnaires baseline, post, and follow-up include the following surveys: PHQ-9, GAD-7, MHC-SF, and Tweets. The individual surveys will be further elaborated in section 2.3. Exclusion within baseline is based on the PHQ-9 and GAD-7 scores. Cutoff point for PHQ-9 and GAD-7 was < 5.

Appendix B

Lists of 27 intervention versions combined by the factors Content, Design, and Feedback

Intervention		*Content*	*Design*	*Feedback*
1.	PNA	Positive Psychology	No gamification elements	Avatar
2.	PNV	Positive Psychology	No gamification elements	Video
3.	PNT	Positive Psychology	No gamification elements	Text
4.	MNA	Meaning Intervention	No gamification elements	Avatar
5.	MNV	Meaning Intervention	No gamification elements	Video
6.	MNT	Meaning Intervention	No gamification elements	Text
7.	CNA	Cognitive Behavioural Therapy	No gamification elements	Avatar
8.	CNV	Cognitive Behavioural Therapy	No gamification elements	Video
9.	CNT	Cognitive Behavioural Therapy	No gamification elements	Text
10.	PCV	Positive Psychology	Competitive gamification	Video
11	MCV	Meaning Intervention	Competitive gamification	Video
12.	CCV	Cognitive Behavioural Therapy	Competitive gamification	Video
13.	MCT	Meaning Intervention	Competitive gamification	Text
14.	PCT	Positive Psychology	Competitive gamification	Text
15.	CCT	Cognitive Behavioural Therapy	Competitive gamification	Text
16.	PCA	Positive Psychology	Competitive gamification	Avatar
17.	MCA	Meaning Intervention	Competitive gamification	Avatar
18.	CCA	Cognitive Behavioural Therapy	Competitive gamification	Avatar
19.	PSV	Positive Psychology	Storyline gamification	Video
20.	MSV	Meaning Intervention	Storyline gamification	Video
21.	CSV	Cognitive Behavioural Therapy	Storyline gamification	Video
22.	PST	Positive Psychology	Storyline gamification	Text
23.	MST	Meaning Intervention	Storyline gamification	Text
24.	CST	Cognitive Behavioural Therapy	Storyline gamification	Text
25.	MSA	Meaning Intervention	Storyline gamification	Avatar
26.	CSA	Cognitive Behavioural Therapy	Storyline gamification	Avatar
27.	PSA	Positive Psychology	Storyline gamification	Avatar

Note. P= Positive Psychology. M= Meaning Intervention. C= Cognitive Behavioural Therapy. N= No gamification elements. C= Competitive gamification. S= Storyline gamification. A= Avatar (feedback). V= Video (feedback). T= Text (feedback)

Appendix C

Adjusted TWEETS items for post questionnaire assessing engagement scores

Item	TWEETS post survey items
1	Using this app did become part of my daily routine
2	The app took me little effort to use
3	I was able to use the app as often as needed (to achieve my goals)
4	This app made it easier for me to work on increasing my wellbeing
5	This app motivated me to increase my wellbeing
6	This app helped me to get more insight into my wellbeing
7	I enjoyed using this app
8	I enjoyed seeing the progress I made in this app
9	This app fits me as a person

Appendix D

Patients Health Questionnaire -9 (PHQ-9)

Item	Over the last 7 days, how often have you been bothered by any of the following problems?	Not at all	Several days	More than half the days	Nearly every day
	Scores	0	1	2	3
1.	Little interest or pleasure in doing things				
2.	Feeling, down, depressed, or hopeless				
3.	Trouble falling or staying asleep, or sleeping too much				
4.	Feeling tired or having little energy				
5.	Poor appetite or overeating				
6.	Feeling bad about yourself- or that you are a failure or have let yourself or your family down				
7.	Trouble concentrating on things, such as reading the newspaper or watching television				
8.	Moving or speaking so slowly that other people could have noticed? Or the opposite- being so fidgety or restless that you have been moving around a lot more than usual				
9.	Thoughts that you would be better off dead or of hurting yourself in some way				

Appendix E

Generalized Anxiety Disorder-7 Questionnaire (GAD-7)

Item	Over the last two weeks, how often have you been bothered by the following problems?	Not at all	Several days	More than half the days	Nearly every day
1.	Feeling nervous, anxious, or on edge	0	1	2	3
2.	Not being able to stop or control worrying	0	1	2	3
3.	Worrying too much about different things	0	1	2	3
4.	Trouble relaxing	0	1	2	3
5.	Being so restless that it is hard to sit still	0	1	2	3
6.	Becoming easily annoyed or irritable	0	1	2	3
7.	Feeling afraid, as if something awful might happen	0	1	2	3