The Impact and Role of Feedback and Engagement in a Digital Health Intervention for Depression

Sebastian J. Stewing
s1990713
s.j.stewing@student.utwente.nl
1st supervisor: dr. Saskia M. Kelders
2nd supervisor: dr. Tessa Dekkers
University of Twente, Enschede

Positive Clinical Psychology & Technology (30 ECs)

Faculty BMS

06/09/2021

Contents

Abstract

Background. People with mental disorders increasingly encounter difficulties to receive adequate treatment and healthcare systems can not sufficiently satisfy the needs of people seeking help. Digital health interventions (DHIs) may help to overcome this discrepancy. However, research shows that patients are oftentimes not fully committed and engaged in DHIs. Different intervention and technological factors (e.g., feedback variants) might positively influence the engagement of DHI users. The aim of this study is to investigate the influence of different feedback categories on both the engagement and depression outcome scores of DHI users as well as to explore whether engagement mediates the relation between feedback categories and depression.

Methods. This study was conducted on a sample of 159 participants who participated in a two-week mobile app intervention with daily exercises derived from evidence-based therapeutic approaches (e.g., CBT). The level of depression was assessed before and after the intervention and engagement scores were measured on day 1, 3, and 7, respectively. ANOVAs were performed to test the main effects from different feedback categories on both engagement and depression. To check for differences between individuals, exploratory analyses were conducted. Mediation analyses were employed to investigate whether engagement mediates the relation between feedback categories and depression.

Results. An overall significant effect of the intervention to reduce depression in the study population was found, $F(1, 156) = 49.18$, $p < .001$, $\eta^2 = .24$. Although on average, no significant differences were found for the influence of different feedback categories on both the engagement and depression outcome scores of DHI users, some individuals strongly deviated from the mean. Furthermore, engagement did not mediate the relationship between different feedback categories and depression outcome scores. Only engagement at T2 predicted post-intervention depression scores and predicted the level of improvement for participants over the course of the intervention ($R^2 = .24$, $F(1, 142)$, $p = .02$).

Conclusion. The study findings suggest that individual participants might benefit from receiving a favourable feedback modality matching their personal needs and preferences. This might positively influence the engagement and outcome scores of DHI users. Future research should investigate factors such as the nature of feedback messages, information architecture, motivation, or using a moderation approach. The present DHI might be used in study populations.

1. Background

In the recent past, it has frequently been reported that people with mental disorders encounter difficulties to receive adequate forms of treatment. Therefore, many of them will remain untreated (Büscher et al., 2020; Karyotaki et al., 2017). In Germany, about 40% of patients who were diagnosed to suffer from a mental disorder after an initial psychotherapeutic assessment had to wait between three and nine months to start psychotherapy in 2019. This translates to an average six months waiting time for psychotherapeutic treatment with numbers expected to further increase due to COVID-19 (Bundespsychotherapeutenkammer, 2021). This may be due to different reasons. On the one hand, the capacities of healthcare systems are being exhausted more often. Overall, the costs to provide sustained health care are not only on a high level already but continue to increase (Karyotaki et al., 2017; Zanaboni et al., 2018). Furthermore, Karyotaki et al. (2017) describe a lack of qualified therapists. Resulting from this, people with mental disorders have limited or poor access to treatment opportunities and will often end up on a waiting list (Büscher et al., 2020; Irish et al., 2020; Zanaboni et al., 2018). On the other hand, it has also been reported that people with mental disorders are hesitant to use traditional forms of treatment. For instance, Josephine et al. (2017) explain that particularly depressed people seem to have a lack of confidence in the healthcare system or might fear being stigmatized (Büscher et al., 2020; Irish et al., 2020; Josephine et al., 2017). However, it appears that they might also avoid approaching treatment opportunities because they either wish to solve the problems themselves or they do not perceive that seeking help is necessary (Büscher et al., 2020; Josephine et al., 2017). Taken together, the aforementioned reasons constitute a range of barriers for people with mental disorders to receive an adequate form of treatment to ultimately alleviate their suffering.

In the last decade, increasingly more attention has been paid to using technological and mobile devices to overcome the barriers of traditional mental healthcare delivery. This approach is commonly referred to as either eMental Health (eMH) or digital health and can best be defined as “mental health services and information delivered or enhanced through the internet or related technologies” (Christensen et al., 2002, p. 3). These services might take the form of digital health interventions (DHIs) presented as different applications via internet- and mobile-based technologies (Josephine et al., 2017). Hereby, DHIs rely on and benefit from the continuously increasing popularity and availability of mobile and digital technologies (Riadi et al., 2020). Liverpool et al. (2020) stress that young people are particularly skilled users of internet and mobile devices who could largely benefit from interventions built on eMH. In addition to that, several health organizations such as the WHO or the United Kingdom’s

National Health Service confirm and support the use of mobile technological devices as suitable tools to provide treatment for different kinds of mental disorders (e.g., depression; Riadi et al., 2020).

The benefits of providing mental health services through technological devices are wide-ranging and may potentially overcome the increasing demands on the healthcare system. In general, DHIs may be used at multiple stages in the treatment of mental disorders. They may help with the early identification and diagnosis of a mental disorder, the overall management, or the analysis or evaluation of the treatment process (Riadi at el., 2020). DHIs may also enhance the availability and accessibility of treatment opportunities. As such, they could grant treatment access to people living in rural and remote areas (Irish et al., 2020, Riadi et al., 2020), mobilize populations avoidant of traditionally delivered mental health interventions (e.g., those in fear of stigmatization; Andrews et al., 2018; Irish et al., 2020; Liverpool et al., 2020; Riadi et al., 2020); or allow large numbers of users to engage in DHIs at any time and from anywhere, thus reducing the increasing costs of healthcare delivery in the long term (Karyotaki et al., 2017; Liverpool et al., 2020; Zanaboni et al., 2018). Through their high accessibility, these interventions may potentially reduce the waiting time to receive a treatment spot (Liverpool et al., 2020). When face-to-face therapy is not readily available (e.g., for people on waitlists), DHIs as a stand-alone treatment option show positive results to reduce, for instance, depressive symptoms (Sethi, 2013). Zanaboni et al. (2018) emphasize that DHIs could even help patients to become more independent in their own health management by offering an increasingly self-directed treatment approach to the users allowing them to track health developments themselves or to support their own informed decision-making (see also Karyotaki et al., 2017; Josephine et al., 2017). In sum, DHIs have great potential to overcome a range of access barriers to traditional forms of mental health treatment delivery.

It has been suggested above that DHIs could be used to treat depression or subthreshold depressive symptoms. In general, positive results have been found for treating depression with different forms of DHIs such as computerized (cCBT) or internet-based (iCBT) cognitive behavioral therapy (Andrews et al., 2018; Liverpool et al., 2020). For instance, Sethi (2013) describes that receiving a self-guided computerized DHI based on CBT principles yielded significant improvements on depression measures as compared to a non-treatment control group that equals a waitlist condition. Although she concluded that DHIs are effective in treating mild to moderate depression, she showed that blended care – the combination of online and face-to-face treatment – was most effective in treating depression overall (Sethi, 2013). In their systematic review and meta-analysis, Josephine et al. (2017) even infer that DHIs can be effective for treating severe depression as well. In addition, they found no significant differences when comparing guided and unguided DHIs which suggests that human contact is not necessarily needed to provide effective treatment using DHIs. For instance, CBT-based DHIs were shown to be promising and effective in reducing depression for populations such as children and young people (Liverpool et al., 2020) or adolescents (Andrews et al., 2018). Furthermore, in their meta-analysis, Karyotaki et al. (2017) found that self-guided CBT-based DHIs can help to reduce the severity of depressive symptoms and lead to a greater treatment response as compared to a waitlist and face-to-face control group. These findings show that DHIs in its different forms can significantly help to disburden the healthcare system and to deliver adequate treatment to everyone in need.

Besides the benefits of technologically driven health interventions, there are also some downsides to consider. Overall, it has been argued that DHIs are not engaging enough to the users or that the full potential of DHIs has not yet been realized (Kelders et al., 2020a; Sharpe et al., 2017). The engagement of DHI users is a commonly investigated issue. However, it lacks a clear definition and conceptualization within the field of eHealth. In general, engagement is described as a multidimensional construct comprising a cognitive, affective and behavioral component (Kelders et al., 2020a). The most comprehensive definition of engagement within the field of eHealth has been proposed by Perski et al. (2017). They specify that the concept of engagement not only includes the extent of DHI usage – reflecting the behavioral component by the amount, frequency, and depth of use – but also entails the subjective experience of the user – describing the cognitive and affective components in terms of their attention, emotions, and interest during use (Perski et al., 2017; Short et al., 2018). To date, the behavioral component has predominantly been focused on (Kelders et al., 2020a). For instance, it has often been assumed that when a DHI is used more often, the positive effects will be greater for the user – a so-called dose-response relationship (Donkin et al., 2011; Kelders et al., 2020a). In recent years, however, researchers became increasingly more aware that engagement with DHIs goes beyond the mere usage of a technological intervention (Kelders et al., 2020a; Perski et al., 2017). Kelders et al. (2020a) question whether these dimensions exhaustively describe the concept of engagement for the field of eHealth, and they theorize whether behavior should also be investigated by the quality of use (e.g., are DHIs used as intended by the designers) or whether negative affect should also play a role in affective engagement (Kelders et al., 2020a). Therefore, further research on engagement and its relation to other concepts is warranted.

To overcome the issues in engagement with DHIs, it has been proposed that choosing a fitting content and design for an intervention may positively influence user engagement

(Kelders et al., 2020a; Sharpe et al., 2017). Sharpe et al. (2017) explain that several factors can influence subsequent engagement with a DHI after use has been initiated. Among these factors are the personalization and tailoring of intervention elements, the ease of set-up and use, tools for self-monitoring as well as including options for feedback and encouragement (Sharpe et al., 2017). They also emphasize that individualized feedback and encouragement in particular may improve the engagement with DHIs (Sharpe et al., 2017; Zagorscak et al., 2020). Yet other research suggests that digital health information (including feedback messages) should be tailored according to the preferences of users (Groeneveld, 2020; Nguyen et al., 2020; Ryanet al., 2018). For instance, Ryan et al. (2018) systematically reviewed the effects of tailoring DHIs to induce weight loss in users. They concluded that a tailored approach is not only viewed more positively by users but also that tailored health information is processed and elaborated upon more deeply (Ryan et al., 2018). Nguyen et al. (2020) confirm these findings. In an experimental study, they provided participants with different modes of information presentation on a website (e.g., text-only, text with visuals, audio-visual, or combinations). They found that tailoring of digital health information according to participantspreferences for information presentation improved the effectiveness of messages and in turn led to increased personal relevance and satisfaction for users when engaging in DHIs (Nguyen et al., 2020). Furthermore, Dekkers et al. (2021) investigated the effects of different design elements on the engagement of DHI users and the effectiveness of DHIs themselves. They found that, for instance, a tunnelled information design - guiding the user through a predetermined sequence of information - was used the longest whereas a matrix design - providing more navigation autonomy to the user - resulted in the highest subjective experience (Dekkers et al., 2021). Lastly, Groeneveld (2020) experimented with differing information variants of feedback messages that were tailored to particular patient profiles - a numerical indication with a brief message, an automated graph, or a message provided by their health careprovider. Overall, most participants were satisfied with their feedback allocation. Nevertheless,only half of their participants reported potential positive effects of the DHI such as reassurance,insight and stimulation by the DHI which indicates that these findings do not apply to everyone (Groeneveld, 2020). Hence, there is an even stronger need to match DHIs with the preferencesof its users.

These studies highlight the importance of tailoring both the content and delivery of digital health information (e.g., feedback messages), show that multiple options for tailoring exist and that increasing the personal relevance of digital health information to DHI users yields positive effects. These findings all line up well with the elaboration likelihood model of persuasion (ELM; Petty & Cacioppo, 1986). Petty and Cacioppo (1986) proposed that as personal relevance increases, people will become increasingly motivated to process information and to elaborate on it, resulting more diligent information processing overall. Applied to the present context, this model might explain the importance of modifying and tailoring the modality of feedback messages of DHIs according to users' needs and preferences to elicit more meaningful, long-lasting and deeper processing of digital health information. Therefore, it appears to be crucial to choose an appropriate modality and a fitting content when providing feedback (Kraft et al., 2017). Tailoring feedback to users' needs and goals has not only been shown to increase personal relevance while working with an intervention (Groeneveld, 2020; Kraft et al., 2017; Nguyen et al., 2020) but might also increase participant engagement and retention with a DHI (Ni Mhurchu et al., 2014; Sharpe et al., 2017). And although different forms of feedback might be equally effective, individual DHI users might be more engaged by different forms of feedback as suggested by Groeneveld (2020).

Recent research has shown that individuals might receive and perceive modified digital health information differently which may affect their engagement with the intervention and ultimately its effectiveness. This study aims to investigate how different modes to deliver feedback within a DHI impact the engagement of users and the effectiveness of DHIs overall. Hereby, the effectiveness will be measured using depression scores. The different modes of feedback used in this study are feedback (1) as a text message, (2) as a text message delivered by a virtual agent, and (3) as a pre-recorded video provided by a human counselor. Research has not yet identified whether one type of feedback is more effective than another. To this end, the following research questions were formulated:

RQ1: *Do different kinds of feedback influence the engagement of digital health intervention users?*

RQ2: *Do different kinds of feedback influence the overall effectiveness of digital health interventions?*

In addition to this, it has been suggested that sustained engagement might result in better outcomes for DHIs. Resulting from this, it was hypothesized that engagement might mediate the relationship between different modes of feedback and the effectiveness of DHIs. Hence, the following research question was formulated:

RQ3: *Does engagement mediate the relationship between different kinds of feedback and the overall effectiveness of digital health interventions?*

2. Methods

2.1 Design

This master thesis is part of a larger study project aimed at developing a personalization approach for eMental Health conducted at the University of Twente in Enschede. The overarching research employs a 3x3x3 full factorial design composed of three variations of selected intervention and technological factors (ITFs), respectively. The three ITFs used in the larger project are 1) the content, 2) feedback variants, and 3) the design of the intervention. For the present study,the focus will solely be on the different forms of feedback in order to investigate their influence on both the engagement of DHI users and the effectiveness of the overall intervention. Participants worked with the intervention for 14 days. Within this time, they completed three engagement measures (1^{st}, 3^{rd}, 7^{th} day of the intervention). Depression was measured before and after the intervention as well as on follow-up measurements after 4 and 8 weeks, respectively. For this study, all three engagement measurements but only the first two depression measurements (pre- & post-intervention) will be used. An overview about the flow of the intervention can be found in Appendix 1a. The study was approved by the Ethics Committee of the Faculty of Behavioral, Management, and Social Sciences at the University of Twente (number: 201118).

2.2 Participants

The original sample population consisted of 770 participants who completed the baseline survey for the study. These participants were older than 18 years of age, showed a general interest in the intervention, were proficient in the English language and possessed a mobile phone. However, participants who – in the baseline survey – had a flourishing mental health according to the Mental Health Continuum – Short Form (MHC-SF; Keyes, 2002) were excluded from the study. In the end, most participants (n = 520) did not complete the post-intervention survey due to the following reasons: they did not start the intervention, they did not register in the corresponding mobile app, or they disengaged from the intervention at some point. In any case, premature dropout resulted in not completing the post-intervention survey which was presented during the last module of the intervention. Hence, only 250 participants completed the post-intervention survey and therefore the whole intervention. Participants occasionally used a different self-generated ID when completing the pre- and post-intervention survey. These had to be adjusted to match one another; the mismatches were dismissed (n = 55). Additionally, a few cases were removed that surprisingly appeared in the post-intervention survey but not in the baseline survey (n = 13). Lastly, another 23 cases were removed because their records for all of the three engagement measurements were missing.

The final study sample consisted of 159 participants of which the majority were female (*f* = 118, 74.2%; *m* = 39, 24.5%; *other* = 2, 1.3%). Their age ranged from 18 to 70 years (*M* = 23.3, *SD* = 8.67); however, most participants were aged between 18 and 22 years (*n* = 121, 76.1%). From the whole sample, 79.2% were students (*n* = 126) whereas only a minority was either working (*n* = 16; 10.1%), unemployed (*N* = 6; 3.8%), retired (*n* = 1; 0.6%), or occupied in another way (*n*= 10, 6.3%). Most participants were German (*n* = 99, 62.3%) but there were also many Dutch participants (*n* = 35, 22%) and some participants from other countries (*n* = 25, 15.7%). In general, no incentives were given for participation, however, students from the University of Twente could enroll for the study through the so- called SONA system and they were granted credits for their participation.

2.3 Materials

2.3.1 Intervention

The present study was conducted via the TIIM app ('the incredible intervention machine'). It is a tool employed by the BMS lab of the University of Twente in Enschede to design and manage digital interventions. This software was used to design the current intervention to increase well-being. In total, 27 different versions of the intervention were constructed based on combinations of selected ITFs from the 3x3x3 research design. These were supposed to have varying effects on the engagement of DHI users and the overall effectiveness of the intervention. Every single intervention version consisted of 14 daily modules that in turn contained one short exercise. These exercises were derived from existing, evidence-based interventions from different therapeutic approaches such as cognitive-behavioral therapy (Merrill et al., 2003; Roth et al., 2004), acceptance and commitment therapy (ACT; Matilla et al., 2016); Powers et al., 2009), and positive psychology (Carr et al., 2020). For instance, in some of the interventions based on positive psychology, participants worked on remembering 'three good things' in which they envisioned and focused on positive experiences that happened during the day. By doing so, positive emotions are fostered and strengthened (Bohlmeijer & Hulsbergen, 2018).

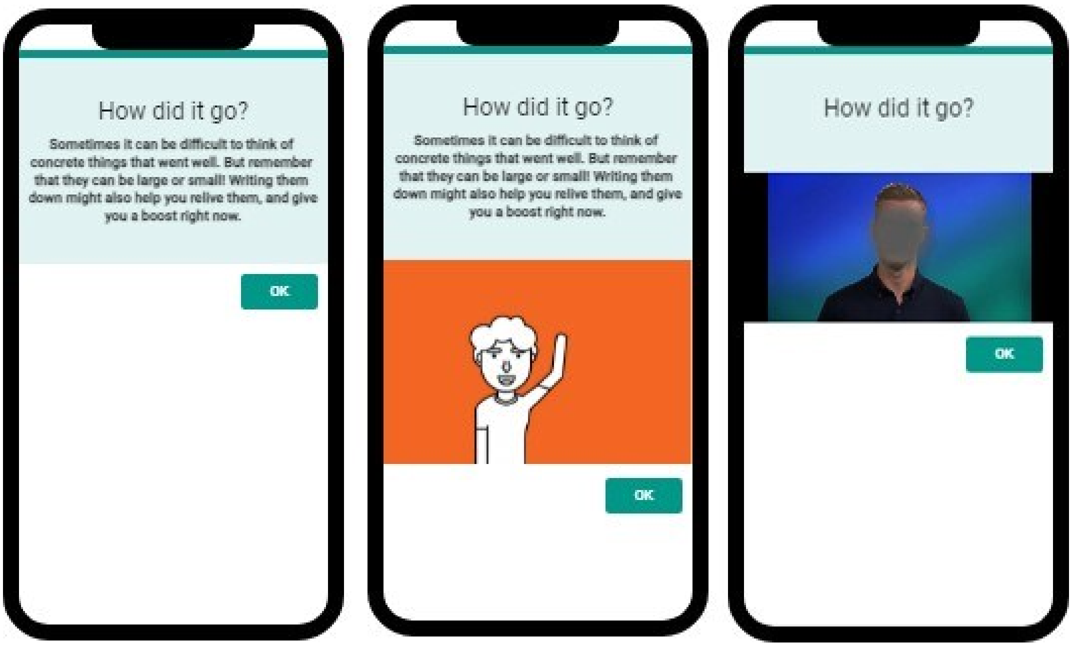
2.3.2 Feedback

For the purpose of this study, the three variations of feedback will be explained more closely. Feedback was provided after having completed the daily exercise. An example of the flow of a daily exercise featuring the feedback provision can be found in Appendix 1b. To allow for reliable comparisons on the varying modality of the feedback messages, the content was always the same between the feedback versions on a particular day. However, the feedback content changed every day to match the exercise at hand. For instance, taken from a version of a positive psychological intervention highlighting the exercise of remembering ‘three good things’, a feedback message for one particular day could read as follows:

> *“How did it go? Sometimes it can be difficult to think of concrete things that went well. But remember that they can be large or small! Writing them down might also help you relive them and give you a boost right now.”*

Figure 1

Examples of varying modalities to provide feedback



Note. From left to right: feedback as (1) a text message, (2) a text message provided by a virtualagent, (3) a pre-recorded video presented by a human counselor.

Examples of varying modalities to provide the feedback messages in the TIIM app can be found in Figure 1. The first version showed the feedback message as a plain written text without any additional features. The second version represents the same written text message as was shown in the first version. This time, however, the text message was accompanied by a virtual agent suggesting that the agent delivers the feedback message. The third version was a pre-recorded video in which a human counselor read out the feedback message. In this version, the written text message was not shown at all so that the user was focusing completely on the spoken words of the counselor.

2.3.3 Measures

This study employed two questionnaires to assess how different forms of feedback influence the engagement with the DHI and the effectiveness of the overall intervention assessed by measures for depression. To measure engagement, the full TWEETS questionnaire was used after the first day, after three days, and after seven days (Kelders et al., 2020b). In total, the TWEETS entails nine items measured on a 5-point Likert scale with possible engagement scores ranging from 9 (*highly engaged*) to 45 (*not engaged*). The subscales consist of three items each that assess behavioral, affective, and cognitive engagement, respectively. The TWEETS was shown to have good psychometric properties (Kelders et al., 2020b). To measure depression, the PHQ-9 questionnaire was used at baseline and to conclude the last day of the intervention (Kroenke & Spitzer, 2002). Its nine items cover the relevant DSM-V criteria needed to diagnose a depressive disorder. In addition to potential preliminary diagnoses of depression, the PHQ-9 can also be used to assess depression severity. The items are measured on a 4-point Likert scale with possible depression scores from 9 (*no depression*) to 36 (*severe depression*). The PHQ-9 was shown to have good psychometric properties (Kroenke et al, 2001).

2.4 Procedure

To recruit participants, convenience and snowball sampling was used via various channels. On the one hand, the researcher consulted his social environment (e.g., family, friends) and social media profiles. On the other hand, the study was uploaded to the SONA system where students can participate in studies conducted by researchers from the University of Twente, Enschede.

Participants were initially contacted via one of the abovementioned channels and an invitation letter (Appendix 2) was provided to brief newly recruited participants about general information about the content, procedure, and theoretical background of the study paired with screenshots of the corresponding application. Within this invitation, they were asked to fill out the baseline survey and to download and enroll in the TIIM app. The baseline survey contained statements asking for participants consent and voluntary participation as well as measures on outcome variables (e.g., depression). After completing the baseline survey, participants were checked against the inclusion and exclusion criteria. Following initial assessment, participants were randomly assigned to one of the 27 intervention types in the TIIM app, and the start of their participation was scheduled for the next day (9 a.m. local time). Participants then worked through the modules of their assigned intervention for 14 consecutive days. Ideally, they worked consistently with the intervention every day and filled in the second survey on the last day. However, it was possible to take longer than 14 days to complete the intervention. In this case, participants' progress was checked regularly, and they were reminded twice to finish both the intervention and the post-intervention survey. At most, participants could take four weeks to complete the intervention. The follow-up surveys were adjusted based on the date of completion and the overall participation ended with finalizing the two follow-up surveys.

2.5 Data analysis

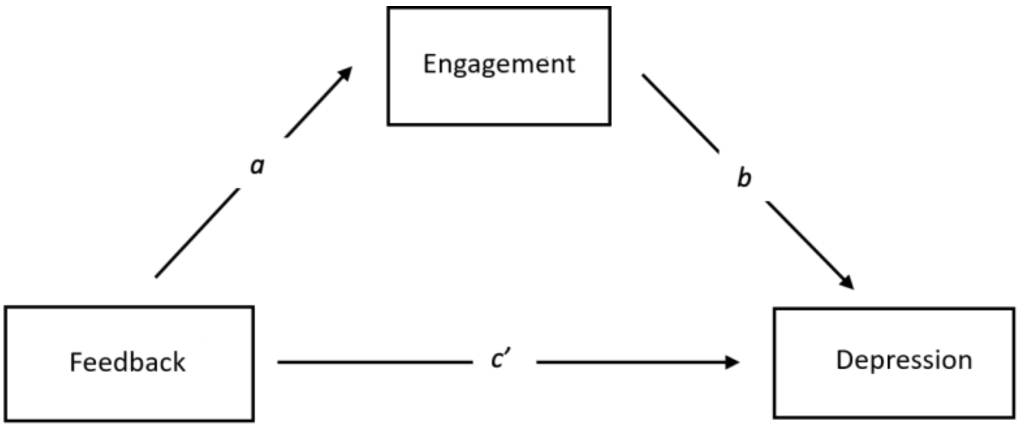
The data was available in five different sources: the pre- and post-intervention surveys containing the depression measures as well as the engagement measures at the first day, the third day, and the seventh day of the intervention. To merge the original data to one final data set, the personal identifier (ID; computed by the users) was used to match participants' data. Email and IP addresses were used as a back-up reference in case participants used a different ID post-intervention as compared to pre-intervention. The personal ID was then matched with another,TIIM-related identifier (TIIM ID), to identify the type of intervention – hence the type of feedback – the participants received. The TIIM ID was also used to merge the data from the engagement measures with the pre- and post-intervention surveys. In the end, all cases were included that contained full responses for the pre- and post-intervention surveys. For the engagement measures, it occurred that responses were occasionally incomplete. These missing data was marked as missing values, but those participants remained in the data set.

To analyze the data, the software *IBM SPSS Statistics* (Version 27) was used. To investigate the relations formulated in research question 1 – the influence of feedback on engagement – and research question 2 – the influence of feedback on depression, an exploratory approach was used mainly drawing from descriptive statistics. These were computed for depression – at T1, T2, and their difference (showing the change of depression over time) – and for engagement – at T1-T3 – both per feedback category and for the whole sample to check for any special occurrences. Furthermore, boxplots were used to check for centrality and spread of the data. A profile plot was used to visualize how average depression scores changed over time. Histograms were also employed to display the distribution of participants for both their change score for depression and engagement at T1 (as one example for engagement). When outliers were observed in the graphical analyses, they were investigated more closely to gain a better understanding for the reasons behind it. To this end, their individual scores were examined manually and additional remarks about personal circumstances (e.g., impact of life events) and experiences with the app (e.g., bugs) were investigated (Appendix 3).

A one-way repeated measures ANOVA was conducted to test the main and interaction effects of the different feedback categories on repeated measures of depression. The assumptions for a one-way repeated measures ANOVA – independent observations, normality, and sphericity – were checked and met. To investigate the main and interaction effects of different feedback categories on engagement, three simple ANOVAs were computed, one for each measurement point of engagement (T1-T3). The assumptions for a simple ANOVA – independent observations, normality, homogeneity – were checked and met. For both types of ANOVAs, it can be expected to have a normal distribution due to the central limit theorem (Field, 2018). Only few outliers have been found, however, these did not represent extreme values. Additionally, ANOVAs are robust against outliers (Field, 2018). For all analyses, the feedback categories were used as a categorical, independent variable and the three measurements of engagement and the depression scores were used as the outcome variables, respectively. The present study also investigates the role of engagement as a mediator for the relation between different feedback categories and depression. The ANOVAs already cover path a and path c' from the mediation model (Figure 2). To also look at path b, three simple linear regressions were computed to check whether engagement at one of the respective measurement points (T1-T3) predicts post-intervention depression scores. The pre-intervention depression scores were also included and controlled for.

Figure 2

Mediation model



To investigate research question 3, a mediation model was used to check whether engagement mediates the relation between different forms of feedback and depression outcome scores. For this purpose, the PROCESS extension version 3.5 for *IBM SPSS Statistics* (Version 27) was used (Hayes, 2017). It was setup with a 95% confidence interval and 5000 bootstrap samples. The latter was used to test indirect effects and to generate a confidence interval around the indirect effect (Field, 2018). PROCESS uses bootstrapping to calculate the mediating effect of engagement on the relationship between different feedback categories and depression outcome scores. For indirect path models (e.g., mediation) the assumption for normal distribution is often questionable (Gajewski et al., 2006). Bootstrapping serves as a robust method for non-normal distributions (Efron, 1979), has a high statistical power, and reduces type 1 errors (Hayes, 2009). The predictor variable (feedback categories) was indicated as a multicategorical variable within the model. Thereby, feedback categories were dummy coded and plotted against each other (1v2, 1v3, 2v3) to investigate the distinct paths in the model. All other options were left on default. The mediation analyses were run three times for each of the three measurement points of engagement, respectively.

3. Results

The final data set included 159 participants. These were almost equally distributed among the different feedback categories. However, participants' characteristics did not differ significantly (p > .05). Table 1 presents more detailed demographic information for the respective feedback categories. Overall, participants were evenly distributedacross the feedback categories for all characteristics. Within each feedback category, significant differences have been found for all respective subcategories for gender (p < .001), age (p < .001) education (p < .001), and nationality (p < .001)

3.1 Descriptive Statistics

Table 2 provides an overview about the average scores and standard deviation for engagement and depression at different measurement points and for all feedback categories including scores for the whole sample. Overall, no significant differences have been found for all measurement points and feedback categories. The mean scores for depression at T2 are slightly lower as compared to depressionat T1 across all feedback categories – as indicated by the difference score. On average, participants tended to show rather strong engagement with the DHI, irrespective of their feedback category. Responses for both depression and engagement almost covered the full range of each scale, respectively, indicating a high variance among participants' responses. For depression, responses ranged from 13 to 33 (4-point Likert scale, 9 items) whereas for engagement, responses ranged from 9 to 42 (5-point Likert scale, 9 items).

Table 1

Demographic information per feedback category

Characteristic	Text		Agent		Video	
	N	%	*n*	%	*n*	%
Participants	50	31.4	57	35.8	52	32.7
Gender						
Male	15	30	15	26.3	9	17.3
Female	35	70	41	71.9	42	80.8
Other	-	-	1	1.8	1	1.9
Age						
< 20	27	54	26	46.8	27	51.3
21-30	15	30	25	45	22	41.8
31-40	3	6	4	7.2	2	3.8
> 40	5	10	2	3.6	1	1.9
Education						
Working	6	12	7	12.6	3	5.8
Student	36	72	45	78.9	45	86.5
Other	8	16	5	9	4	7.6
Nationality						
German	33	66	33	75.4	33	63.5
Dutch	12	24	10	17.5	13	25
Other	5	10	14	24.6	6	11.5

Table 2

Mean scores and standard deviations at different measurement points and divided by feedback category

	Text (n = 50)		Agent (n = 57)		Video (n = 52)		Total (N = 159)	
	M	SD	M	SD	M	SD	M	SD
Depression – T1	20.42	4.47	20.98	4.14	20.31	5.04	20.58	4.54
Depression – T2	17.02	4.6	18.56	4.88	18.02	4.45	17.9	4.67
Engagement – T1	20.6	4.34	21.04	6.45	19.25	4.43	20.31	5.25
Engagement – T2[a]	21	4.4	21.46	5.51	20.4	4.3	20.99	4.81
Engagement – T3[b]	20.16	4.64	21.2	5.74	21.02	5.79	20.83	5.42
Depression – Difference Score	-3.4	5.17	-2.42	4.07	-2.29	5.31	-2.69	4.85

Note. [a] engagement T2: $N = 145$, text: $n = 46$, agent: $n = 54$, video: $n = 45$; [b] engagement T3: $N = 143$, text: $n = 44$, agent: $n = 55$, video: $n = 44$.

$p > .05$ for all measurement points and for all feedback categories.

3.1.2 Depression

Figure 2.

Averaged Change of Depression Scores from Pre- to Post-Intervention Displayed per Feedback Category

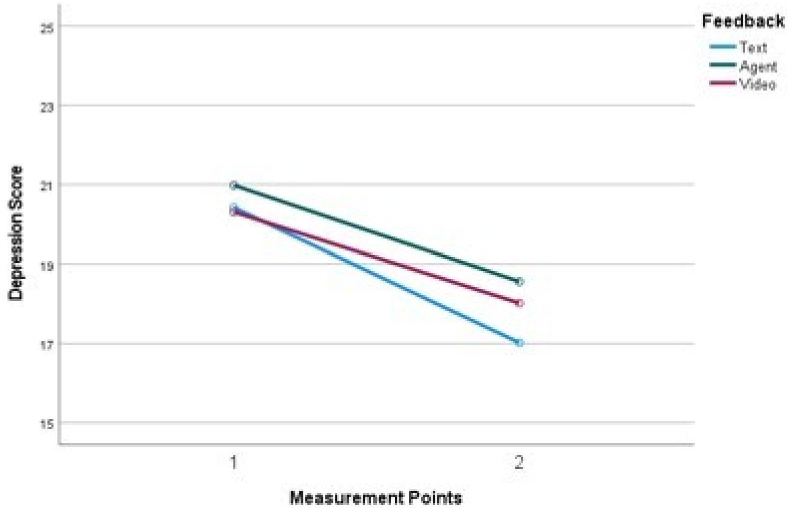


Table 2 displays similar baseline scores for depression at T1 across all categories and there were no special occurrences when checking with a boxplot (Appendix 4a). The post-intervention measurement for depression at T2 showed almost similar descriptive statistics. However, a boxplot revealed that the data for the feedback category 'agent' is skewed towards the top resembling a larger variance for participants with high scores on depression. Also, outliers were found for every feedback category indicating that some participants are highly depressed at T2 (Appendix 4b). In comparison, a high variance was shown for the change of depression over time for the whole sample (Appendix 4c). The interquartile range and median were both lower in T2 as compared to T1. Outliers exist towards both the upper and lower end of the boxplot. For some participants, the depression has worsened a lot whereas for others, it has improved a lot. Small improvements were also found for every category with slightly better improvements in the 'text' category (Appendix 4d). Overall, change – positive and negative – in depression scores was almost similar across categories ($p > .05$), but change was smallest for the 'agent' category as indicated by a shorter interquartile range. Figure 2 visually compares the depression scores from T1 and T2 per feedback category resembling the change of depression over time.

3.1.2 Engagement

The average engagement at T1 was very similar for all feedback categories (Table 2). However, the range of responses was larger for the 'agent' category with a maximum score of 42. This finding was supported by another boxplot (Appendix 4e) that showed similar shapes for the 'text' and 'agent' categories except that the latter category had 3 outliers with high negative engagement scores. Engagement within the 'video' category was different. The interquartile range was shorter overall meaning that the individual cases were more similar. However, the 'video' category also had one outlier with a high positive score and 3 outliers with high negative scores. It appears that overall, participants showed similar responses but that a few individual responses strongly deviated from the majority.

3.2 Inferential Statistics

3.2.1 Research question 1

No statistically significant main effects of different feedback categories on engagement were found. This concerned all three simple ANOVAs for the respective engagement measures. For engagement T1, no significant main effect of the different feedback categories was found, $F(2, 156) = 1.697$, $p = .19$, $\eta^2 = .02$. For engagement T2, no significant main effect of the different feedback categories was found, $F(2, 142) = .597$, $p = .55$, $\eta^2 = .01$. For engagement T3, no statistically significant main effect of the different feedback categories was found, $F(2, 140) = .489$, $p = .61$, $\eta^2 = .01$

3.2.2 Research question 2

A one-way repeated measures ANOVA demonstrated a statistically significant main effect of the intervention on a change in depression scores, $F(1, 156) = 49.18$, $p < .001$, $\eta^2 = .24$. Given the descriptive statistics that were computed before (see chapter 3.1), depression scores consistently improved across the different feedback categories. However, no statistically significant main effect of the different feedback categories on depression scores was found, $F(2, 156) = 9.427$, $p = .45$, $\eta^2 = .01$.

Three simple linear regression analyses – investigating whether the different engagement measures (T1-T3) can predict depression scores at T2 – led to the following results: engagement at T1 did not predict depression scores at T2, $R^2 = .22$, $F(1, 156) = 3.15$, $p = .08$; engagement at T2 predicted depression scores at T2, $R^2 = .24$, $F(1, 142)$, $p = .02$., so 24% of the variance in depression scores at T2 was predicted from engagement at T2; engagement at T3 also predicted depression scores at T2, $R^2 = .268$, $F(1, 140) = 7.73$, $p = .01$, therefore showing a significant relationship in which 26.8% of the variance in depression scores at T2 was predicted from engagement at T3. The results show that engagement measures increasingly predicted the variance in depression outcome scores at T2 depending on how close the engagement measures were to the post-intervention depression measurement.

3.2.3 Research question 3

In general, no mediating effect of engagement was found for each measurement point, respectively. Figure 5 provides an overview of the mediation models for each measurement point of engagement (Figure 5a-5c).

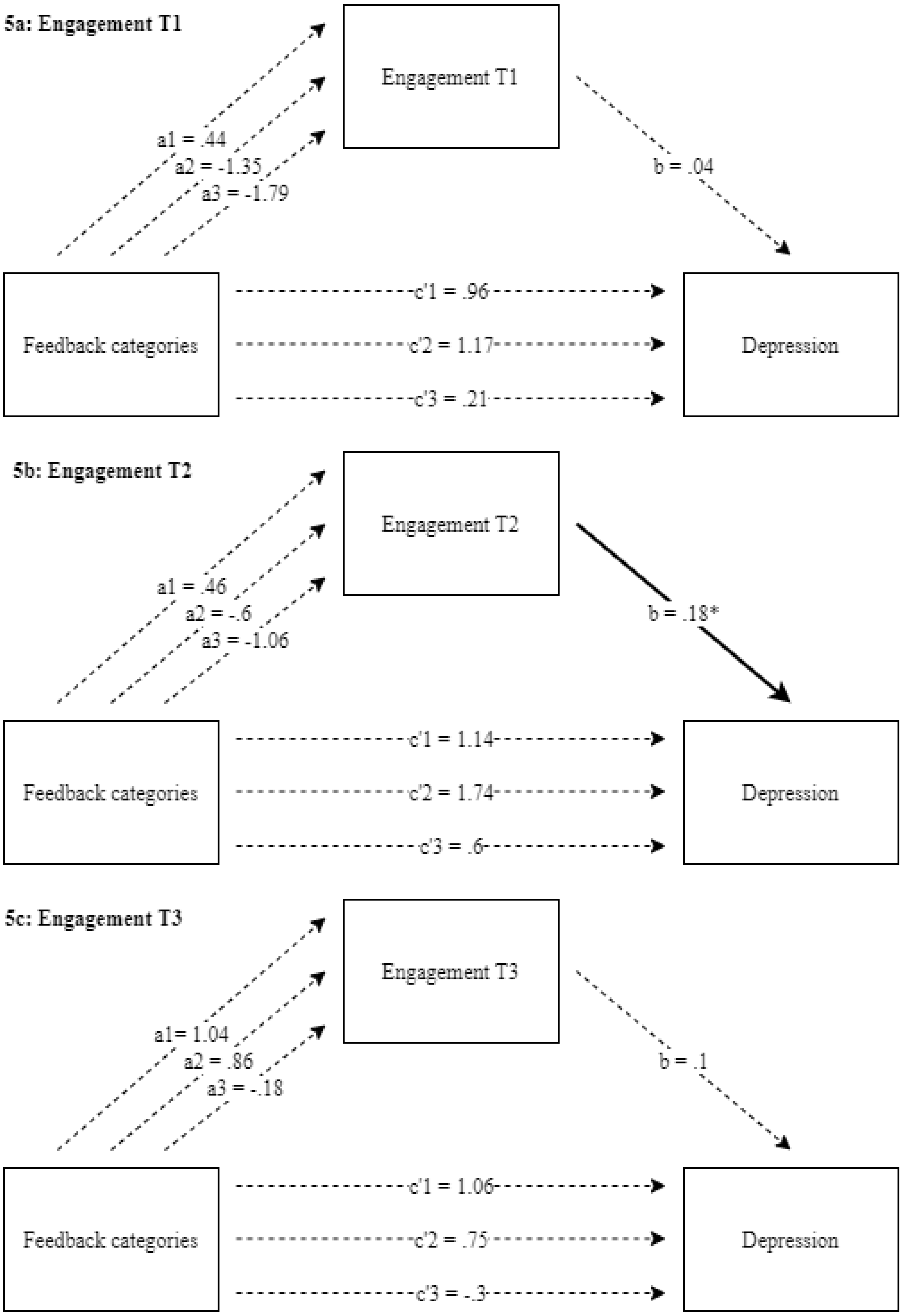
Engagement at T1. Results indicated that the different feedback categories are not indirectly related to depression outcome scores through their relationship with engagement at T1 (Figure 5a). The distinct paths were all nonsignificant and had no predictive value. A 95% confidence interval based on 5000 bootstrap samples confirmed that the indirect effect included zero for all feedback categories (text vs. agent: [-.20, .26]; text vs. video: [-.45, .16]; agent vs. video: [-.53, .23]).

Engagement at T2. Results indicated that the different feedback categories are not indirectly related to depression outcome scores through their relationship with engagement at T2 (Figure 5b). The distinct paths were all nonsignificant except for that engagement at T2 predicted depression outcome scores at T2, $R^2 = .05$, $F(3, 141) = 2.73$, $p = .025$. All other paths had no predictive value. A 95% confidence interval based on 5000 bootstrap samples confirmed that the indirect effect included zero for all categories (text vs. agent: [-.31, .53]; text vs. video: [-.56, .27]; agent vs. video: [-.73, .15]).

Engagement at T3. Results indicated that the different feedback categories are not indirectly related to depression outcome scores through their relationship with engagement at T3 (Figure 5c). The distinct paths were all nonsignificant and had no predictive value. A 95% confidence interval based on 5000 bootstrap samples confirmed that the indirect effect included zero for all feedback categories (text vs. agent: [-.14, .48]; text vs. video: [-.15, .54]; agent vs. video: [-.36, .34]).

Figure 5

Mediation Models for each Measurement Point of Engagement



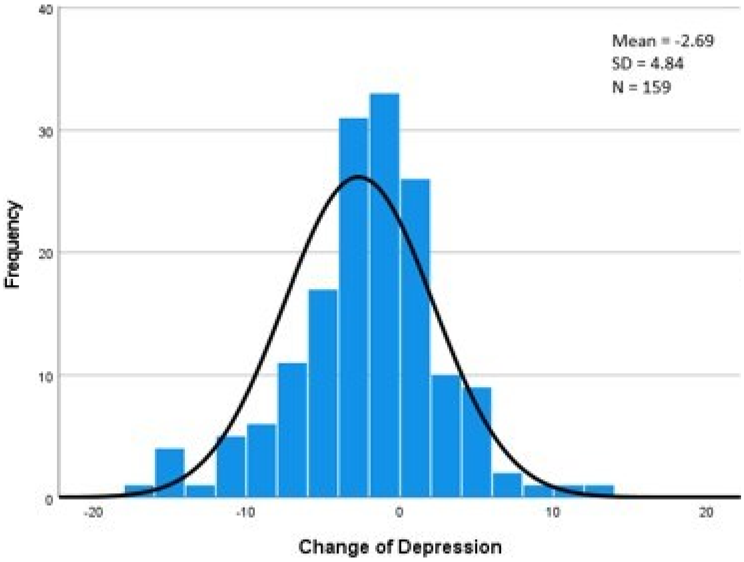
Note. These mediation models predict depression scores from different feedback categories with a mediating effect of engagement at different measurement points (T1-T3). Statistics are unstandardized regression coefficients. Dotted lines represent nonsignificant relations; bold lines represent significant relations. The different paths for a and c' represent several dummy-coded comparisons for the feedback categories: a1/c'1 = text vs. agent; a2/c'2 = textvs. video; a3/ c'3 = agent vs. video.

3.3 Outliers

Some participants strongly deviated from the mean both in the change of their depression scores over time and their engagement scores at T1 (see Chapter 3.1). Detailed information and additional remarks about their individual scores are displayed in Appendix 3. Regarding the change of depression scores over time, six participants strongly improved, and two participants strongly worsened as compared to the mean (Figure 3). As for the engagement scores at T1, outliers existed who – despite rather high engagement scores overall – had particularly low engagement scores (Figure 4). This concerned six participants (see also Appendix 4e).

Figure 3.

Distribution of Participants for Change of Depression Scores Over Time



Because strong deviations in responses were discovered, the sample was split across feedback categories and then assigned to subcategories of the respective scales for both depression (T1 and T2) and engagement (T1-3). Table 3 shows the distribution of depression scores across feedback categories. It reveals that the majority of participants had a moderate level of depression at T1 whereas the majority of participants had a low level of depression at T2, highlighting overall improvements in depression scores which applies to every feedback category. Table 4 shows the distribution of engagement scores across feedback categories. Here, the engagement scores in the ‘agent’ category were particularly low as compared to the other two feedback categories.

Figure 4.

Distribution of Participants for Engagement at T1

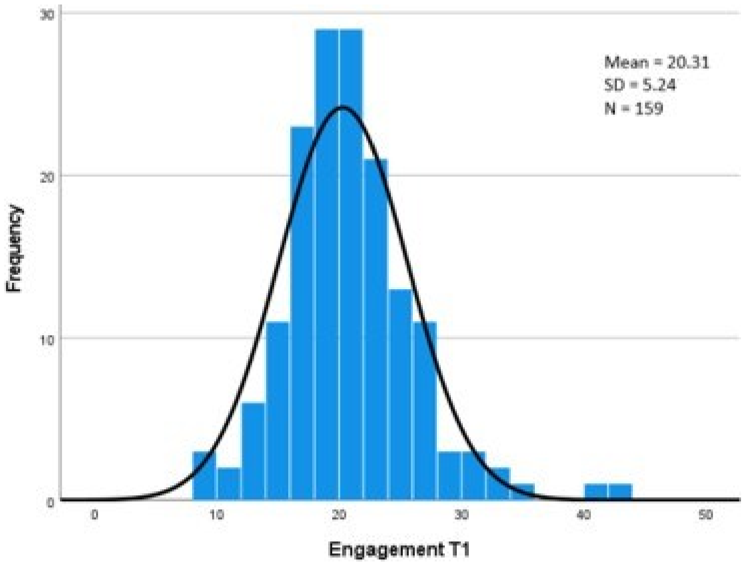


Table 3

Distribution of depression scores for participants across the feedback categories

Depression levels	Text (*n* = 50)		Agent (*n* = 57)		Video (*n* = 52)	
	N	%	*n*	%	*n*	%
			Depression at T1			
low	15	30	14	25.2	16	30.8
moderate	29	58	38	68.4	29	55.7
high	6	12	5	9	7	13.5
			Depression at T2			
low	32	64	28	50.4	27	51.9
moderate	15	30	25	45	21	40.4
high	3	6	4	7.2	4	7.7

Note. The depression levels are derived from the PHQ-9 scale: low = 9-17, moderate = 18-26, high = 27-36.

Table 4

Distribution of engagement scores for participants across the feedback categories

Engagement levels	Text (*n* = 50)		Agent (*n* = 57)		Video (*n* = 52)	
	N	%	*n*	%	*n*	%
			Engagement at T1			
very low	-	-	2	3.5	-	-
low	5	10	7	12.3	4	7.6
high	33	66	32	56.1	30	59.6
very high	12	24	16	28.1	17	32.7
missing	-	-	-	-	-	-
			Engagement at T2			
very low	-	-	2	3.8	-	-
low	5	10.9	7	13.3	3	5.7
high	31	67.4	34	62.9	31	58.9
very high	8	21.7	11	20.4	9	17.1
missing	4	8	3	5.3	7	13.5
			Engagement at T3			
very low	-	-	1	1.8	-	-
low	4	9.1	10	18	7	13.3
high	25	54.5	32	58.2	26	49.4
very high	16	36.4	12	21.8	11	20.9
missing	6	12	2	3.5	8	15.4

Note. The engagement scores are derived from the TWEETS: very low = 36-45, low = 27-35, high = 18-26, very high = 9-17.

4. Discussion

This study was set out to investigate two objectives. On the one hand, it aimed to examine how different modes to deliver feedback within DHIs impact both the engagement of users (RQ1) and the effectiveness of DHIs overall (RQ2). It was hypothesized that different modes of feedback will show no difference on average, but that some variation in engagement and effectiveness might be expected for individual participants. On the other hand, it was explored if engagement would mediate the relationship between different modes of feedback and the effectiveness of DHIs (RQ3).

The findings of the present study revealed no differences on average for the different feedback categories as was expected. First, the different feedback categories did not influence engagement at its different measurement points (T1-T3). Further exploratory analyses – investigating differences for individual participants – found no significant differences on average between the different feedback categories, however, for some participants, deviations from the mean were discovered. Second, the different feedback categories did not influence depression outcome scores despite overall improvements in depression scores indicating an effect of the intervention. On average, there were no significant differences between the different feedback categories, but for some individuals, the intervention resulted in either a high or low change of depression scores over the two-week intervention period. Third, no mediating effect of engagement on the relationship between different feedback categories and depression outcome scores were found. However, regression analyses revealed that engagement at certain measurement points during the intervention (T2 and T3) predicted depression outcome scores.

Ultimately, the first two research questions both aimed to investigate how different modes of feedback could be used to develop a more engaging experience for DHI users and thus to design more effective interventions. The results of the present study showed that different feedback categories did neither significantly influence the engagement of DHI users nor the effectiveness of the overall DHI although large individual differences in engagement and effectiveness were discovered. These exploratory findings indicate that while the mode of feedback does not influence engagement directly, personalizing the delivery of feedback might have positive effects for individual DHI users. Participants were randomly assigned to one feedback modality and, therefore, did not receive a tailored feedback allocation. However, some participants experienced very positive effects which indicates that these individuals were most likely assigned to a feedback category that represented a particularly good match to their personal needs and preferences. These findings are in line with previous research highlighting the value of tailoring and personalizing digital health information to DHI users. These studies concluded that a tailored approach might elicit reassurance, insight, and stimulation by the DHI (Groeneveld, 2020), might increase personal relevance and satisfaction when engaging with a DHI (Nguyenet al., 2020), lead to deeper processing and elaboration of the presented health information (Ryan et al., 2018), or increase participant engagement and retention with a DHI (Ni Mhurchuet al., 2014; Sharpe et al., 2017). Evidence in support of a tailored approach to deliver digital health information exists. Previous research has highlighted the importance of feedback to improve engagement in DHIs (Sharpe et al., 2017; Zagorscak et al., 2020). In contrast to this, the present study could not show that different feedback categories influenced both engagement and depression scores, although some individuals deviated from the mean. These nonsignificant effects could be explained by the randomized assignment of participants to their respective feedback category which possibly resulted in unfitting matches. To investigate a tailored approach more accurately, a similar design could be employed as was used in the Groeneveld study where the feedback assignment was based on previously developed patient profiles (Groeneveld, 2020). However, this would require additional research preceding the feedback assignment to discover which participant profiles match the respective feedback categories.

Future research should not only alter feedback modalities but rather consider other factors as well to increase the impact of feedback on the engagement of DHI users and ultimately its effectiveness. For instance, in a rapid review on DHIs for weight management, Sharpe et al. (2017) described that participants disliked the generic nature of feedback and regarded it as being impersonal and repetitive. Although this intervention implemented increasingly personified versions of feedback – from text to video delivered by a human counselor – the provision of feedback was rather static and generic, providing only one version of a feedback message to all participants but differently framed. Perhaps, feedback messages could be designed to refer back to earlier input from the users themselves instead of presenting generic and automated feedback messages. Furthermore, the feedback messages were integrated in a predetermined sequence that users were obliged to follow in their daily work with the intervention – a so-called tunneled information design. Dekkers et al. (2021) suggested that a tunneled information design was used the longest whereas an alternative matrix design – providing navigation autonomy to the user – resulted in the highest subjective experience for DHI users. Sieverink et al. (2017) also assumed that successful participation in DHIs is not necessarily determined by the length of usage but more by the usefulness of DHI elements to reach individual goals and needs. Changing the overall structural design of the intervention – where feedback messages are placed – and granting users more control over decision-making and navigation – how and when feedback messages are accessible – might therefore increase the impact of feedback by increasing the subjective experience for individual users – which is a central component in the concept of engagement in the field of eHealth (Kelders et al., 2020a; Perski et al., 2017). In support of this elaboration, a study on health technology engagement identified several determinants of user engagement with DHIs such as the overall satisfaction with, or the navigability and the ease of use of DHIs (Cole-Lewis et al., 2019; see also Short etal., 2015). Adapting the nature, placement, and availability of feedback messages more closely to users' actual needs and goals might therefore result in a higher level of perceived autonomy for the users. This might in turn lead to more personal relevance and ownership of the DHI and – according to the ELM (Petty & Cacioppo, 1986) – elicit more meaningful, deeper, and longer processing of the feedback messages as well as increased engagement in the long run. Arnold et al. (2020) also found that complex navigation and perceiving a resources (e.g., DHIs) as irrelevant or useless to oneself leads to limited use of the resource itself.

Another facet of feedback that might be interesting to investigate in the future – and that might influence the engagement of DHI users – is the regularity and continuity of feedback. In the present study, participants continuously received extrinsic feedback presented in different ways. It can be argued that for some participants, the regularity of external feedback might have been too high. According to the self-determination theory (Deci & Ryan, 2012), high levels of external feedback might reduce the intrinsic motivation and persistence of those participants that are genuinely intrinsically motivated and engaged, leading to lower engagement with the DHI. This was confirmed by a study investigating predictors for the engagement with a self-guided online intervention for psychosis (Arnold et al., 2019). They showed that high levels of external influence on participants predicted lower engagement. Consequently, tailoring and personalizing DHIs to the preferences of users could lead to more identification with DHIs and in turn lead to higher levels of intrinsic motivation, persistence, and engagement with DHIs. Future research could further investigate this in two ways. First, the two-week intervention period could be split in two parts. During the first week, participants would receive their allocated feedback category as was implemented in the present study. In the second week, participants could be offered an option to turn the feedback messages on and off based on their preferences. In this way, researchers could not only explore how certain levels of engagement during the first week affect participants' feedback behavior during the second week of the intervention but also to investigate the degree to which participants make use of feedback messages. Second, researchers could present all feedback categories to the participants on a daily basis. This might provide hints about how engagement scores and feedback behavior develop over the course of the whole intervention period.

The third research question investigated whether engagement mediates the relationship between different feedback categories and the overall effectiveness of DHIs. The results showed that no mediating effect of engagement on all three measurement points (T1-T3) was found. So far, a mediating role of engagement for the relationship between intervention and technological factors (ITFs) and mental health outcomes has not explicitly been researched yet for the field of eHealth. However, as an exemplary ITF, feedback – and variants of it – has been shown to positively affect the engagement of DHI users (Nguyen et al., 2020; Sharpe et al., 2017; Zagorscak et al., 2020). Furthermore, research on different DHIs has shown that user engagement may positively influence (mental) health outcomes such as social anxiety (Rice et al., 2020), weight management (Sharpe et al., 2017) or depression and anxiety (Graham et al., 2020; Karyotaki et al., 2017; Wu et al., 2021). Therefore, evidence exists for the relationships of the different paths in the proposed mediation model of this study. Feedback has been shown to affect user engagement and engagement in turn positively influences health outcomes. The findings of the regression analyses indicate that engagement measures at the third and seventh day of the intervention were predictive of post-intervention depression scores. However, in the mediation analyses, this was only confirmed for engagement at third day of the intervention (T2), but not for engagement at the seventh day (T3). The different findings could be explained by the use of different independent variables in the respective analyses. For the regression analyses, the post-intervention scores were used which means that the engagement measures were only related to one point in time. For the mediation analyses, however, the change score of depression – the difference from pre- to post-intervention – was used which resembles developments over the course of the whole intervention. This might explain why the engagement measurement at the seventh day of the intervention influenced the post-intervention depression scores, but not the level of improvement over the course of the whole intervention. However, both analyses found that engagement at the third day of the intervention predicted post-intervention depression scores and influenced the level of improvement for participants. Perhaps, participants' engagement culminated at the third day of the intervention influencing the remainder of the intervention. Future research could build on this to gain more insights about when and how engagement – and developments over time – can best be measured. And although no mediating effect of engagement has explicitly been found in the present study, evidence from the literature – and in parts from this study – suggests that relations between the employed variables exist. Another possibility is that a mediation approach for engagement was an inappropriate design to investigate differences between individual participants or the different feedback categories. The random assignment of participants to the different feedback categories

– regardless of their actual preferences – might have prevented finding a mediating effect of engagement. Since a good match was found for only a few participants, it is possible that a mediating effect might have disappeared. Using a moderation approach for engagement might therefore be more applicable to explore individual differences because it might reveal differences in the strength of engagement on the relationship between different feedback categories and depression outcome scores. For instance, future research could investigate how inter- and intrapersonal changes in engagement scores produce different depression outcome scores. These insights could help to design 'reverse profiles' in which the information about successful feedback allocation might be used to match certain participant characteristics. Ultimately, future studies could use this to tailor the feedback categories to participants which might enhance the overall effectiveness of DHIs.

4.1 Strengths and limitations

Two aspects of this study could count as both a strength and a limitation. The first aspect was that the participants were randomly assigned to one of the intervention types. On the one hand, this allowed for unbiased assignments to the different intervention types. On the other hand, randomization might have been a hindering factor to research a personalization approach. Future research could try to combine both of these approaches to answer the question whether the different feedback categories affect engagement and outcomes on an individual level. For instance, in an experimental group, participants would be given the opportunity to select their favourite choice from the different intervention and technological factors (e.g., feedback) employed in this study ('the personalization group'). The control group would be randomly assigned to one of the intervention types as was implemented in the present study ('the randomization group'). In the end, these groups could be compared to one another, and this could generate hints towards participants' needs and preferences (see also Nguyen et al., 2020). The second aspect concerns the composition of the sample. Participants were mostly female, of young age, currently studying, and from Germany. On the one hand, this could count as a strength because inferences could be made about this particular population. On the other hand, this could be a limitation because the findings cannot be generalized to a greater population with more variety in demographic characteristics.

One strength of this study were the hard exclusion criteria. This allowed for a clear cut off to people with flourishing mental health and only those participants were included in the final sample that had languishing or moderate mental health. This is important because the overall design of the intervention significantly improved depression scores which means that such interventions might help people with languishing to moderate mental health to reduce depressive symptoms. Another strength was to integrate several measurement points to assess the engagement of participants during the intervention. This was valuable because it allowed to investigate developments of engagement over time. However, it remained unclear at what time point it is best to measure participants' engagement and future research is advised to delve into this.

Next to the aforementioned strengths, there were also some limitations for the present study. For instance, the research design was suboptimal because no control group was included in the study. This could be improved by randomly assigning participants to intervention types that do not contain any kind of feedback. Another limitation is that throughout data collection, technical issues with the TIIM app impeded error-free participation for some users during the intervention period (e.g., login issues, delays in progress, modules were repeatedly presented to users). This could have had a negative impact on the engagement of the affected participants. The biggest limitation, however, is the fact that mostly those people were included in the final sample who were to some degree engaged anyway. For instance, 520 participants were excluded from the analyses because they did not complete the post-intervention survey. These participants might still have participated in the intervention for some time to finish some or all of the engagement measurements during the first week of the intervention. For future research, it might be interesting to include the engagement scores of these participants as well to investigate their level of engagement before dropping out of the intervention.

4.2 Implications for research and practice

This study generated a range of valuable insight for future research projects to further investigate the interplay of engagement with health outcomes and intervention and technological factors in DHIs. For instance, when researching feedback provision within DHIs, not only the modality could be altered, but the feedback could also be designed less generic and repetitive and in turn impact user engagement on a more personal basis. This might be realized by asking participants for their personal preferences in the baseline survey of the intervention. Furthermore, future projects could try to change the information architecture towards a matrix design (Dekkers et al., 2021) because granting more autonomy and control to the user yields a higher subjective experience for the users which might positively affect their engagement. Furthermore, the feedback provision could be adjusted in a few other ways. First, feedback could be designed to appeal more to the intrinsic motivation of the participants by integrating, for instance, particularly positive feedback that connects to prior exercises, successes, or even failures (Burgers et al., 2015). Second, the feedback could be personalized by including a reference in the feedback messages that refers to earlier engagement measures. For instance, participants could be encouraged when they have completed the daily exercise despite low levels of engagement. Third, the amount of feedback could be adjusted to match participants needs, for instance, by either providing participants with a toggle option for feedback during the second week or by allowing participants to choose from all three feedback categories every day. Fourth, participants' level of intrinsic motivation could be controlled for in the baseline survey as well. This study did not find any significant effects of the different feedback categories on both engagement and depression. It might be that participants received too much external feedback which might have reduced the intrinsic motivation and persistence for users and in turn reduced their levels of engagement with a DHI (Arnold et al., 2019; Deci & Ryan, 2012). Lastly, the sample of the present study and its findings were limited to characteristics of a certain population – female, young age, students, and German. The study could be replicated in larger and more diverse populations as well to be able to generalize the study findings.

The last implication for research is also relevant for the practical implications of the present study. Because the overall design of the intervention was found to be effective in reducing depression scores, this intervention might be used to assist student populations during times of languishing or moderate mental health. Study counselors could recommend this app to affected students after an initial meeting. Especially during times of the COVID-19 pandemic where in-person contact is limited, this intervention could be helpful for students to improve their wellbeing and to reduce mental health issues.

5. Conclusion

This study generated valuable insights into the impact and role of feedback and engagement in a digital health intervention for depression. Although on average, no significant differences have been found for the influence of different feedback categories on both the engagement of DHI users and depression outcome scores, some individual users appeared to highly benefit from a specific feedback modality that matched their personal needs and preferences. These findings provided valuable hints to increase the engagement of DHI users by further tailoring and personalizing feedback variants. Future research is advised to pursue further investigation into factors such as the nature of feedback messages (generic vs. personalized), information architecture (tunneled vs. matrix design), motivation (extrinsic vs. intrinsic), or using a moderation approach for engagement. This DHI might also be relevant for practical use in student populations with languishing to moderate mental health because it could help improve their well-being or reduce mental health issues.

6. References

Andrews, G., Basu, A., Cuijpers, P., Craske, M., McEvoy, P., English, C., & Newby, J. (2018). Computer Therapy for the Anxiety and Depression Disorders is Effective, Acceptable and Practical Health Care: An Updated Meta-analysis. *Journal of Anxiety Disorders*, *55*, 70–78. https://doi.org/10.1016/j.janxdis.2018.01.001

Arnold, C., Villagonzalo, K. A., Meyer, D., Farhall, J., Foley, F., Kyrios, M., & Thomas, N. (2019). Predicting Engagement with an Online Psychosocial Intervention for Psychosis: Exploring Individual- and Intervention-level Predictors. *Internet Interventions*, *18*, 100266. https://doi.org/10.1016/j.invent.2019.100266

Arnold, C., Williams, A., & Thomas, N. (2020). Engaging With a Web-Based Psychosocial Intervention for Psychosis: Qualitative Study of User Experiences. *JMIR Mental Health*, *7*(6), e16730. https://doi.org/10.2196/16730

Bohlmeijer, E., & Hulsbergen, M. (2018). *Using Positive Psychology Every Day: Learning How to Flourish* (1st ed.). Routledge.

Bunderpsychotherapeutenkammer (2021, March 29). *BPtK-Auswertung: Monatelange Wartezeiten bei Psychotherapeut*innen*. BPTK. https://www.bptk.de/bptk-auswertung-monatelange-wartezeiten-bei-psychotherapeutinnen/

Burgers, C., Eden, A., van Engelenburg, M. D., & Buningh, S. (2015). How Feedback Boosts Motivation and Play in a Brain-training Game. *Computers in Human Behavior*, *48*, 94–103. https://doi.org/10.1016/j.chb.2015.01.038

Büscher, R., Torok, M., Terhorst, Y., & Sander, L. (2020). Internet-Based Cognitive Behavioral Therapy to Reduce Suicidal Ideation. *JAMA Network Open*, *3*(4), e203933. https://doi.org/10.1001/jamanetworkopen.2020.3933

Carr, A., Cullen, K., Keeney, C., Canning, C., Mooney, O., Chinseallaigh, E., & O'Dowd, A. (2020). Effectiveness of Positive Psychology Interventions: A Systematic Review and Meta-analysis. *The Journal of Positive Psychology*, 1–21. https://doi.org/10.1080/17439760.2020.1818807

Christensen, H., Griffiths, K. M., Evans, K., Australia. Department of Health and Ageing, National Mental Health Strategy, Australian National University. Centre for Mental Health Research, National Mental Health Information Strategy Committee, Australian Health Ministers' Advisory Council. National Mental Health Working Group. Information Strategy Committee, & Australia. Mental Health and Special Programs Branch. (2002). *E-Mental Health in Australia*. Commonwealth Department of Health and Ageing.

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. https://doi.org/10.2196/14052

Deci, E. L., & Ryan, R. M. (2012). Self-determination Theory. In P. A. M. Van Lange, A. W.Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 416–436). Sage Publications Ltd. https://doi.org/10.4135/9781446249215.n21

Dekkers, T., Melles, M., Vehmeijer, S. B. W., & de Ridder, H. (2021). Effects of Information Architecture on the Effectiveness and User Experience of Web-Based Patient Education in Middle-Aged and Older Adults: Online Randomized Experiment. *Journal of Medical Internet Research*, *23*(3), e15846. https://doi.org/10.2196/15846

Donkin, L., Christensen, H., Naismith, S. L., Neal, B., Hickie, I. B., & Glozier, N. (2011). A Systematic Review of the Impact of Adherence on the Effectiveness of e-Therapies. *Journal of Medical Internet Research*, *13*(3), e52. https://doi.org/10.2196/jmir.1772

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, *7*(1). https://doi.org/10.1214/aos/1176344552

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

Gajewski, B. J., Lee, R., Thompson, S., Dunton, N., Becker, A., & Wells, V. (2006). Non-normal Path Analysis in the Presence of Measurement Error and Missing Data: A Bayesian Analysis of Nursing Homes' Structure and Outcomes. *Statistics in Medicine*, *25*(21), 3632–3647. https://doi.org/10.1002/sim.2478

Graham, A. K., Greene, C. J., Kwasny, M. J., Kaiser, S. M., Lieponis, P., Powell, T., & Mohr, D. C. (2020). Coached Mobile App Platform for the Treatment of Depression and Anxiety Among Primary Care Patients. *JAMA Psychiatry*, *77*(9), 906. https://doi.org/10.1001/jamapsychiatry.2020.1011

Groeneveld, B.S. (2020). *Talking the walk: Applying Data-driven Patient Profiles in the Design of Tailored Services in Orthopaedics* (Doctoral dissertation). https://doi.org/10.4233/uuid:7c866494-52b6-4a53-baff-f6b2443ef13

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium. *Communication Monographs*, *76*(4), 408–420. https://doi:10.1080/03637750903310360

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.

Irish, M., Zeiler, M., Kuso, S., Musiat, P., Potterton, R., Wagner, G., Karwautz, A., Waldherr, K., & Schmidt, U. (2020). Students’ Perceptions of an Online Mental Health Intervention: A Qualitative Interview Study. *Neuropsychiatrie*. Published. https://doi.org/10.1007/s40211-020-00383-5

Josephine, K., Josefine, L., Philipp, D., David, E., & Harald, B. (2017). Internet- and Mobile-based Depression Interventions for People with Diagnosed Depression: A Systematic Review and Meta-analysis. *Journal of Affective Disorders*, *223*, 28–40. https://doi.org/10.1016/j.jad.2017.07.021

Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., Mackinnon, A., Meyer, B., Botella, C., Littlewood, E., Andersson, G., Christensen, H., Klein, J. P., Schröder, J., Bretón-López, J., Scheider, J., Griffiths, K., Farrer, L., Huibers, M. J. H., . . . Cuijpers, P. (2017). Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms. *JAMA Psychiatry*, *74*(4), 351. https://doi.org/10.1001/jamapsychiatry.2017.0044

Kelders, S. M., Bohlmeijer, E. T., Pots, W. T., & van Gemert-Pijnen, J. E. (2015). Comparing Human and Automated Support for Depression: Fractional Factorial Randomized Controlled Trial. *Behaviour Research and Therapy*, *72*, 72–80. https://doi.org/10.1016/j.brat.2015.06.014

Kelders, S. M., van Zyl, L. E., & Ludden, G. D. S. (2020a). The Concept and Components of Engagement in Different Domains Applied to eHealth: A Systematic Scoping Review. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00926

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. https://doi.org/10.2196/17757

Keyes, C. L. M. (2002). The Mental Health Continuum: From Languishing to Flourishing in Life. *Journal of Health and Social Behavior*, *43*(2), 207. https://doi.org/10.2307/3090197

Kraft, S., Wolf, M., Klein, T., Becker, T., Bauer, S., & Puschner, B. (2017). Text Message Feedback to Support Mindfulness Practice in People with Depressive Symptoms: A Pilot Randomized Controlled Trial. *JMIR MHealth and UHealth*, *5*(5), e59. https://doi.org/10.2196/mhealth.7095

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatric Annals*, *32*(9), 509–515. https://doi.org/10.3928/0048-5713-20020901-06

Liverpool, S., Mota, C. P., Sales, C. M. D., ČUš, A., Carletto, S., Hancheva, C., Sousa, S., Cerón, S. C., Moreno-Peral, P., Pietrabissa, G., Moltrecht, B., Ulberg, R., Ferreira, N., & Edbrooke-Childs, J. (2020). Engaging Children and Young People in Digital Mental Health Interventions: Systematic Review of Modes of Delivery, Facilitators, and Barriers. *Journal of Medical Internet Research*, *22*(6), e16317. https://doi.org/10.2196/16317

Lyden, J. R., Zickmund, S. L., Bhargava, T. D., Bryce, C. L., Conroy, M. B., Fischer, G. S., Hess, R., Simkin-Silverman, L. R., & McTigue, K. M. (2013). Implementing Health Information Technology in a Patient-Centered Manner: Patient Experiences with an Online Evidence-Based Lifestyle Intervention. *Journal for Healthcare Quality*, *35*(5), 47–57. https://doi.org/10.1111/jhq.12026

Mattila, E., Lappalainen, R., Välkkynen, P., Sairanen, E., Lappalainen, P., Karhunen, L., Peuhkuri, K., Korpela, R., Kolehmainen, M., & Ermes, M. (2016). Usage and Dose Response of a Mobile Acceptance and Commitment Therapy App: Secondary Analysis of the Intervention Arm of a Randomized Controlled Trial. *JMIR MHealth and UHealth*, *4*(3), e90. https://doi.org/10.2196/mhealth.5241

Mclaughlin, M., Delaney, T., Hall, A., Byaruhanga, J., Mackie, P., Grady, A., Reilly, K., Campbell, E., Sutherland, R., Wiggers, J., & Wolfenden, L. (2021). Associations Between Digital Health Intervention Engagement, Physical Activity, and Sedentary Behavior: Systematic Review and Meta-analysis. *Journal of Medical Internet Research*, *23*(2), e23180. https://doi.org/10.2196/23180

Merrill, K. A., Tolbert, V. E., & Wade, W. A. (2003). Effectiveness of Cognitive Therapy for Depression in a Community Mental Health Center: A Benchmarking Study. *Journal of Consulting and Clinical Psychology*, *71*(2), 404–409. https://doi.org/10.1037/0022- 006x.71.2.404

Nguyen, M. H., Bol, N., & King, A. J. (2020). Customization versus Personalisation of Digital Health Information. *European Journal of Health Communication*, *1*(1), 30–54. https://doi.org/10.47368/ejhc.2020.003

Ni Mhurchu, C., Whittaker, R., McRobbie, H., Ball, K., Crawford, D., Michie, J., Jiang, Y., Maddison, R., Waterlander, W., & Myers, K. (2014). Feasibility, Acceptability and Potential Effectiveness of a Mobile Health (mHealth) Weight Management Programme for New Zealand Adults. *BMC Obesity*, *1*(1). https://doi.org/10.1186/2052-9538-1-10

O’Brien, H. L., & Toms, E. G. (2009). The Development and Evaluation of a Survey to Measure User Engagement. *Journal of the American Society for Information Scienceand Technology*, *61*(1), 50–69. https://doi.org/10.1002/asi.21229

Perski, O., Blandford, A., West, R., & Michie, S. (2016). 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

Petty, R.E., Cacioppo, J.T. (1986). The Elaboration Likelihood Model of Persuasion. Advances in Experimental Social Psychology, 19(1), 124-205.

Powers, M. B., zum Vörde Sive Vörding, M. B., & Emmelkamp, P. M. (2009). Acceptance and Commitment Therapy: A Meta-Analytic Review. *Psychotherapy and Psychosomatics*, *78*(2), 73–80. https://doi.org/10.1159/000190790

Riadi, I., Kervin, L., Teo, K., Churchill, R., & Cosco, T. D. (2020). Digital Interventions for Depression and Anxiety in Older Adults: Protocol for a Systematic Review. *JMIR Research Protocols*, *9*(12), e22738. https://doi.org/10.2196/22738

Rice, S., O’Bree, B., Wilson, M., McEnery, C., Lim, M. H., Hamilton, M., Gleeson, J., Bendall, S., D’Alfonso, S., Russon, P., Valentine, L., Cagliarini, D., Howell, S., Miles, C., Pearson, M., Nicholls, L., Garland, N., Mullen, E., McGorry, P. D., & Alvarez-Jimenez, M. (2020). Leveraging the Social Network for Treatment of Social Anxiety: Pilot Study of a Youth-specific Digital Intervention with a Focus on Engagement of Young Men. *Internet Interventions*, *20*, 100323. https://doi.org/10.1016/j.invent.2020.100323

Roth, A., Fonagy, P., Parry, G., Target, M., & Woods, R. (2004). *What Works for Whom?, Second Edition: A Critical Review of Psychotherapy Research* (2nd ed.). The Guilford Press.

Ryan, K., Dockray, S., & Linehan, C. (2019). A Systematic Review of Tailored eHealth Interventions for Weight Loss. *DIGITAL HEALTH*, *5*, 205520761982668. https://doi.org/10.1177/2055207619826685

Sethi, S. (2013). Treating Youth Depression and Anxiety: A Randomised Controlled Trial Examining the Efficacy of Computerised versus Face-to-face Cognitive Behaviour Therapy. *Australian Psychologist*, *48*(4), 249–257. https://doi.org/10.1111/ap.12006

Sharpe, E. E., Karasouli, E., & Meyer, C. (2017). Examining Factors of Engagement with Digital Interventions for Weight Management: Rapid Review. *JMIR Research Protocols*, *6*(10), e205. https://doi.org/10.2196/resprot.6059

Short, C.E., Rebar, A.L., Plotnikoff, R.C., Vandelanotte, C. (2015). Designing Engaging Online Behaviour Change Interventions: A Proposed Model of User Engagement. The European Health Psychologist, 17(1), 32-38.

Short, C. E., DeSmet, A., Woods, C., Williams, S. L., Maher, C., Middelweerd, A., Müller, A. M., Wark, P. A., Vandelanotte, C., Poppe, L., Hingle, M. D., & Crutzen, R. (2018). Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies. *Journal of Medical Internet Research*, *20*(11), e292. https://doi.org/10.2196/jmir.9397

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

Tang, J., Abraham, C., Stamp, E., & Greaves, C. (2014). 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

Wu, A., Scult, M. A., Barnes, E. D., Betancourt, J. A., Falk, A., & Gunning, F. M. (2021). Smartphone Apps for Depression and Anxiety: A Systematic Review and Meta-analysis of Techniques to Increase Engagement. *Npj Digital Medicine*, *4*(1). https://doi.org/10.1038/s41746-021-00386-8

Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., Merchant, G. C., Naughton, F., & 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

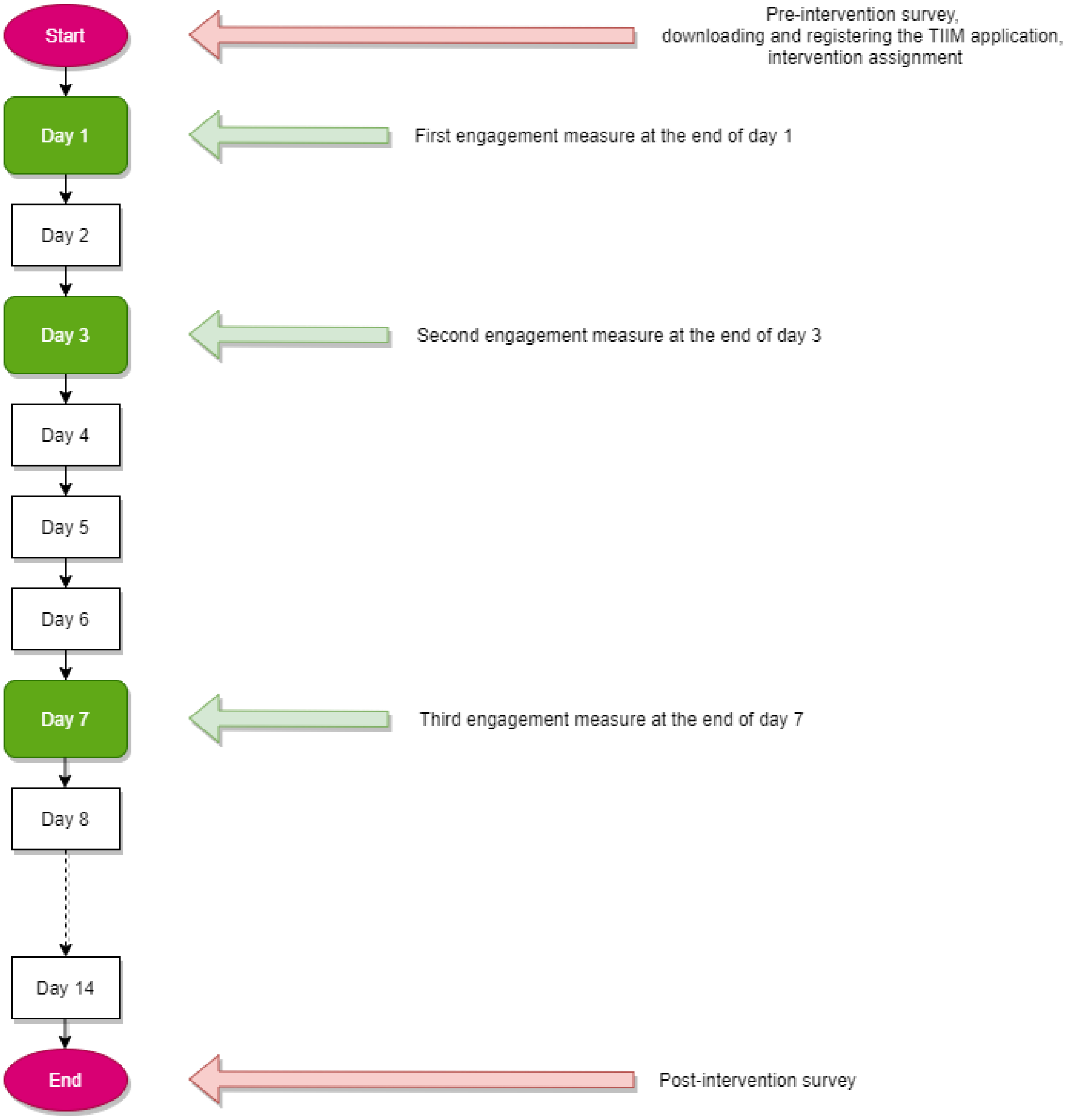
Zagorscak, P., Heinrich, M., Schulze, J., Böttcher, J., & Knaevelsrud, C. (2020). Factors Contributing to Symptom Change in Standardized and Individualized Internet-based Interventions for Depression: A Randomized-controlled Trial. *Psychotherapy*, *57*(2),237–251. https://doi.org/10.1037/pst0000276

Zanaboni, P., Ngangue, P., Mbemba, G. I. C., Schopf, T. R., Bergmo, T. S., & Gagnon, M. P. (2018). Methods to Evaluate the Effects of Internet-Based Digital Health Interventions for Citizens: Systematic Review of Reviews. *Journal of Medical Internet Research*, *20*(6), e10202. https://doi.org/10.2196/10202

7. Appendices

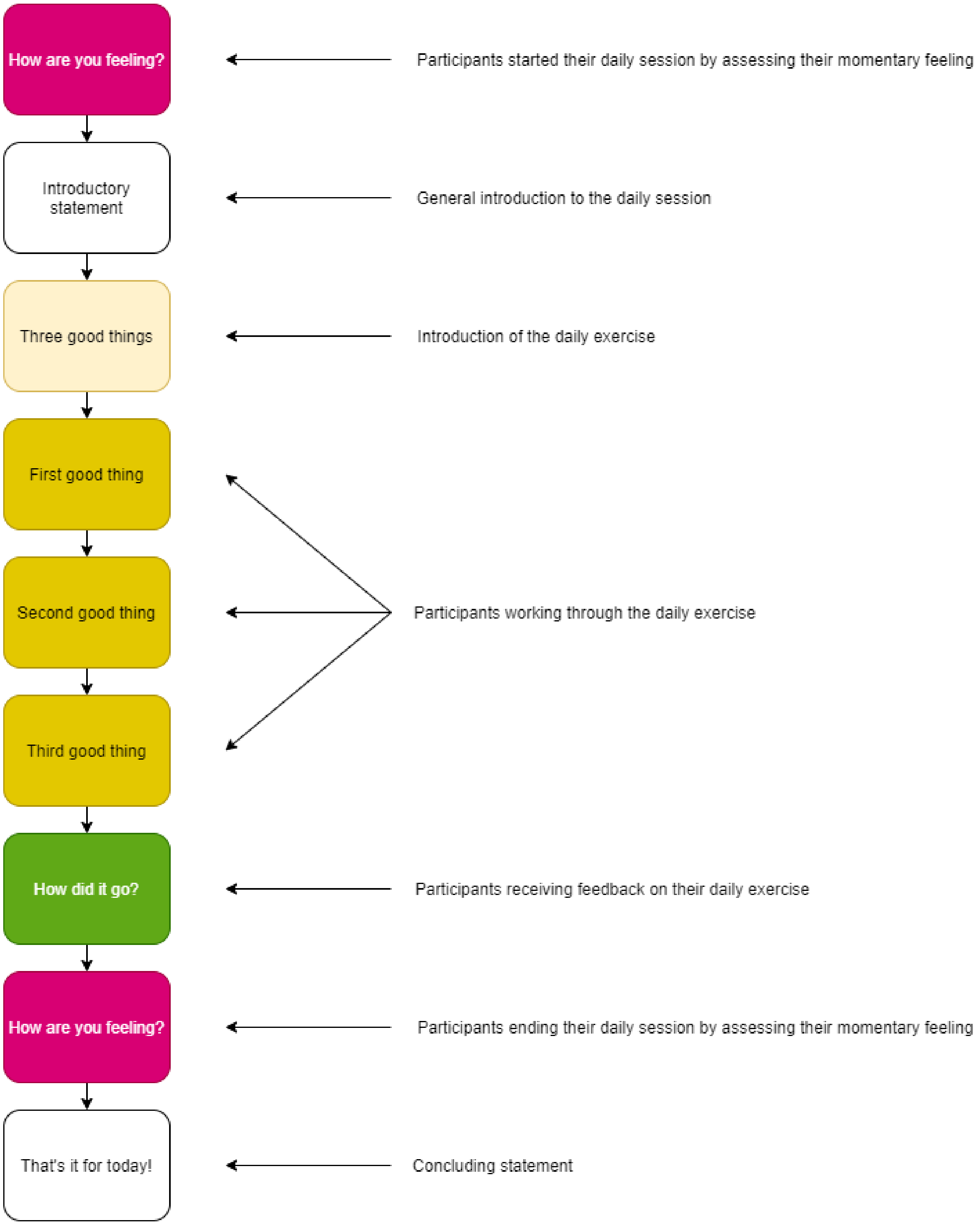
Appendix 1a

Flow chart of the 14-Day Intervention Period highlighting the Different Measurement Points for Depression and Engagement



Appendix 1b

Flowchart of the First Day taken from one Intervention Type featuring a Positive Psychological intervention and the Exercise of Three Good Things



Appendix 2

Invitation Letter Used for the Recruitment of Participants

Study invitation: Learn to flourish with an app!

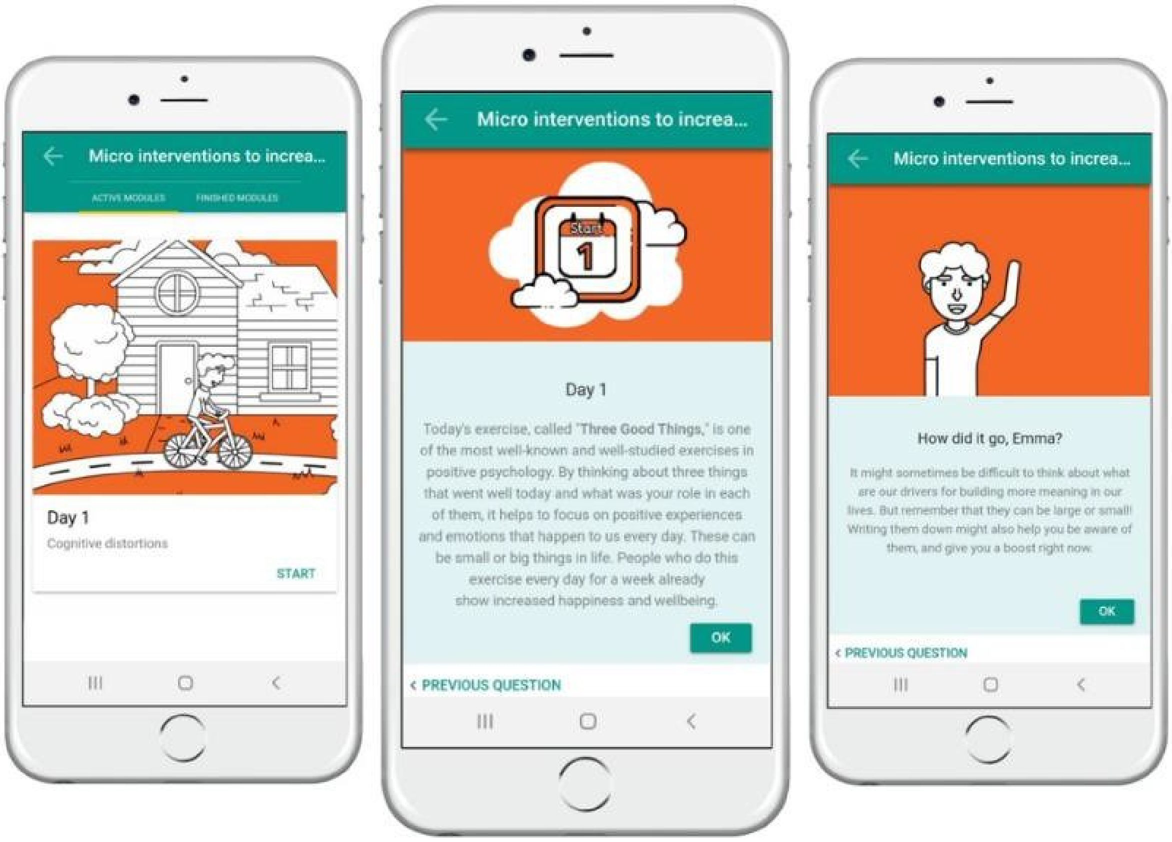
Dear,

We would like your help with a research study. A group from the University of Twente is doing a research study that **aims to improve your wellbeing using a mobile app**.

It is **100% online** and in **English**.

In this study, you will:

- Use a mobile app for **2 weeks**, every day for a couple of minutes;
- Each day, you will get a **short exercise** in the morning that you can complete during the day;
- These exercises are based on **well-known, evidence-based interventions** on *Positive Psychology*, *Cognitive Behavioural Therapy*, or *Meaning and Purpose*;
- The information that you provide in answering the exercises will be kept private;
- Your **daily mood and engagement** with the intervention will be measured via short questionnaires and used anonymously for research purposes;
- You will be invited to participate in **four short online surveys** with questions about how you are feeling at that point in time.



How will it work? Step-by-step

- First, you fill out a [baseline survey] with some brief information about you and about how you are feeling;
- After that, **enroll and install the app**: the instructions for this will be given as soon as you finish the previous step;
- With the app installed, you can **start the exercises**. It will take 2 weeks to complete;
- We will invite you to fill out **another three short surveys** in 2, 4, and 8 weeks.

To participate in this study, CLICK HERE.

You must be 18 years old or more and able to install and use a mobile app for two weeks. Your participation is anonymous and voluntary. You can withdraw at any time, for any reason.

Contact Us

Do you have any questions or concerns about this study?

Please, feel free to contact us.

Appendix 3

Additional Remarks and Information about the Observed Outliers

<table>
<tr><th rowspan="2">Participants (feedback category)</th><th colspan="3">Depression scores</th><th colspan="3">Engagement scores</th><th rowspan="2">Additional remarks</th></tr>
<tr><th>Pre</th><th>Post</th><th>Difference</th><th>T1</th><th>T2</th><th>T3</th></tr>
<tr><td>48 (text)</td><td>19</td><td>31</td><td>12</td><td>9</td><td>9</td><td>18</td><td>Consistent, very high engagement (avg. scores: T1/T2 = 1; T3 = 2); highest neg. change in whole sample</td></tr>
<tr><td colspan="8">Neg. impact[1]: work insecurity
App-pos[2]: easy to use and interactive
App-neg[3]: potentially video content would be good to see case</td></tr>
<tr><td>19 (text)</td><td>28</td><td>13</td><td>-15</td><td>16</td><td>15</td><td>16</td><td>Consistent, high engagement; depression strongly improved</td></tr>
<tr><td colspan="8">App positive: Info and rationale behind intervention, practical, not too time consuming, providing new perspective</td></tr>
<tr><td>36 (text)</td><td>31</td><td>15</td><td>-16</td><td>20</td><td>19</td><td>24</td><td>Moderate engagement, trending to high engagement; depression strongly improved</td></tr>
<tr><td colspan="8">Neg impact: The extraction of all of the teeth in my lower jar.
Bug explanation[4]: The selection of language says Nederlands but the questions remain in English. I could not proceed to the last survey due to an unknown password. In one of the questions about values appeared - blank - instead of the answers I gave the days before. It is supposed to be a reminder. When answering the email I would not get an answer or the person was no longer at Twente University.
App-pos: The daily reminder to live after your values and the XY which shows progress.
App-neg: I would sometimes forget it. So one more reminder?</td></tr>
<tr><td>53 (agent)</td><td>24</td><td>12</td><td>-12</td><td>16</td><td>19</td><td>11</td><td>Consistent, high engagement; depression strongly improved</td></tr>
<tr><td colspan="8">App-pos: It is easy to work with, the design was very cool! I was looking forward to see new designs. I liked that there was a “person” guiding me, speaking to me using my name.
App-neg: The app never reminded me in the evening if I hadn’t done a task. And if I started to fill something in, it was not saved, and I had to fill it in again later.</td></tr>
<tr><td>54 (agent)</td><td>29</td><td>14</td><td>-15</td><td>22</td><td>22</td><td>19</td><td>Moderate engagement; depression strongly improved</td></tr>
<tr><td colspan="8">Neg. impact: Death of a close family member.
Bug explanation: In the beginning it was not possible to use it. And in the end, I wasn't reminded of using it, which made it difficult to remember to use it.
App-pos: Easy language. The "person" who speaks with me looks happy and is encouraging
App-neg: More variety within the tasks; the tasks were often repeated. Maybe more tasks that one can do immediately, so one does not forget to do it. Or another reminder in the end of the day.</td></tr>
</table>

Continued.

<table>
<tr><th rowspan="2">Participants (feedback category)</th><th colspan="3">Depression</th><th colspan="3">Engagement</th><th rowspan="2">Additional Remarks</th></tr>
<tr><th>Pre</th><th>Post</th><th>Difference</th><th>T1</th><th>T2</th><th>T3</th></tr>
<tr><td>105 (agent)</td><td>23</td><td>31</td><td>8</td><td>21</td><td>21</td><td>22</td><td>Consistent, moderate engagement; depression worsened</td></tr>
<tr><td colspan="8"><u>Neg. impact</u>: Weaning off medication
<u>App-pos:</u> It allows you to answer questions however you want, the answers are for yourself, so you don't feel judged or pressure to answer a particular way
<u>App-neg</u>: I have an issue where I swipe away notifications without looking at them, I'm not sure how the app could change to prevent that but perhaps different forms of notifications or recommending participants set daily alarms if notifications don't work well for them</td></tr>
<tr><td>117 (video)</td><td>31</td><td>14</td><td>-17</td><td>31</td><td>-</td><td>31</td><td>Low engagement, engagement T2 skipped; highest pos. change in depression</td></tr>
<tr><td colspan="8"><u>App-pos</u>: it makes you reflect on positive things and reminds you of good moments in life that are worth living for
<u>App-neg</u>: I am not sure what one could improve</td></tr>
<tr><td>130 (video)</td><td>31</td><td>16</td><td>-15</td><td>24</td><td>22</td><td>20</td><td>Moderate engagement; depression strongly improved</td></tr>
<tr><td colspan="8"><u>App-pos:</u> Idea itself, micro lectures at end of unit.
<u>App-neg:</u> Hard to integrate into schedule, too much cognitive effort to manage.</td></tr>
<tr><td>52 (agent)</td><td>22</td><td>22</td><td>0</td><td>41</td><td>36</td><td>38</td><td>Outlier for all 3 engagement measurements; consistent, very low engagement; no change in depression</td></tr>
<tr><td colspan="8"><u>Neg-impact</u>: Living. My mother calling and accusing me of making the ministry of finance checking her accounts, potentially leading to prison time for her
<u>App-pos</u>: Good for people who are directionless, I guess.
<u>App-neg</u>: It helps those who are confused, directionless or immature and feel down and no one else. It's essentially completely meaningless because I know my values and goals.</td></tr>
<tr><td>65 (agent)</td><td>28</td><td>24</td><td>-4</td><td>35</td><td>28</td><td>26</td><td>Consistent, low engagement; depression minimally improved</td></tr>
<tr><td colspan="8"><u>Bug-explanation</u>: Often, the promised notifications were not visible to me. They were inconsistent which is why I had to set an alarm to not miss a day.
<u>App-pos</u>: It encourages you to think about positive things at least once a day, even if it's just for a few minutes.
<u>App-neg:</u> I'm not sure. Instead of having so many different activities I think it might be beneficial to have fewer, especially because some of them need some practice.</td></tr>
</table>

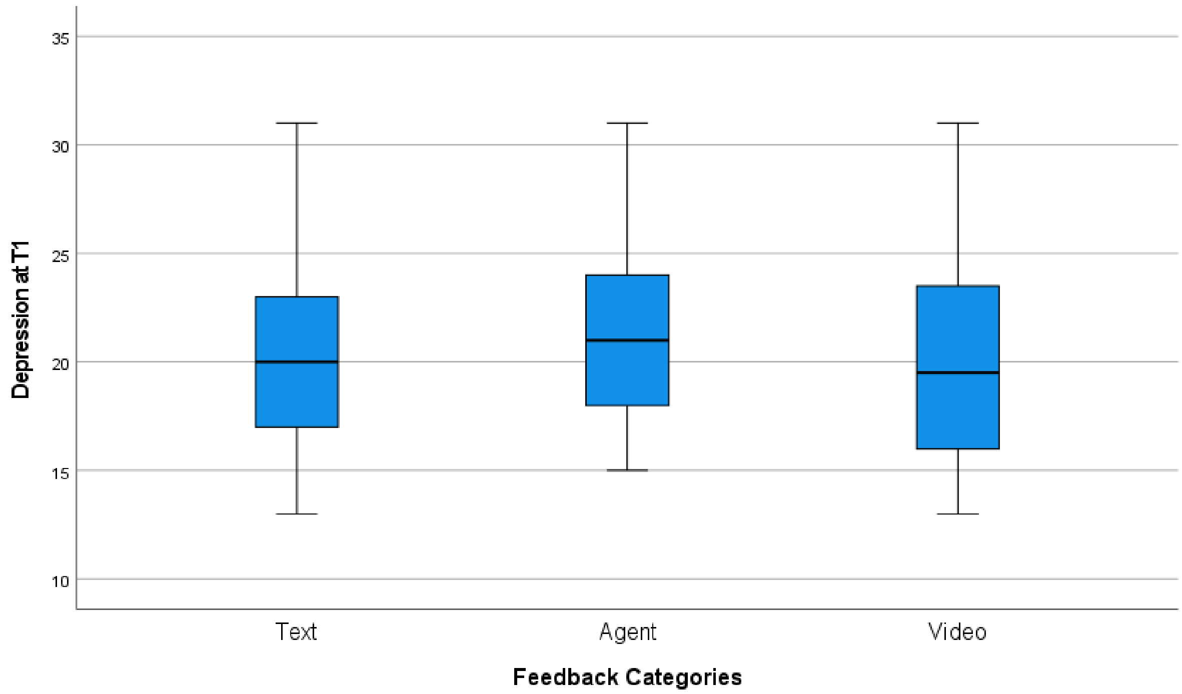
Continued.

<table>
<tr><th rowspan="2">Participants (feedback category)</th><th colspan="3">Depression</th><th colspan="3">Engagement</th><th rowspan="2">Additional Remarks</th></tr>
<tr><th>Pre</th><th>Post</th><th>Difference</th><th>T1</th><th>T2</th><th>T3</th></tr>
<tr><td>69 (agent)</td><td>29</td><td>33</td><td>4</td><td>42</td><td>39</td><td>33</td><td>Extreme value for engagement T2; consistent, very low engagement; depression got minimally worse</td></tr>
<tr><td colspan="8"><u>Neg-impact</u>: loss of multiple family members (parent and 3 grandparents), ending a romantic relationship, failing multiple exams in university
<u>App-pos:</u> clearly structured, nice to look at
<u>App-neg:</u> it was really nice :) no complaints here</td></tr>
<tr><td>112 (video)</td><td>25</td><td>19</td><td>-6</td><td>9</td><td>18</td><td>17</td><td>Consistent, very high engagement (avg. scores: T1 = 1, T2 = 2); depression minimally improved</td></tr>
<tr><td colspan="8"><u>App-pos:</u> It makes you think about things you would normally not think about regularly
<u>App-neg</u>: It could maybe give some advice every day, like a little thing one could do to calm down or something that makes you appreciate the small, nice things in life or that helps you relax a bit</td></tr>
<tr><td>143 (video)</td><td>24</td><td>15</td><td>-9</td><td>30</td><td>-</td><td>30</td><td>Low engagement, T2 skipped; depression improved</td></tr>
<tr><td colspan="8"><u>Bug-explanation:</u> It said I will receive a notification for the next day, but I did not receive one. And even if you go out of an exercise for example if you are being called or you want to google a word from the app, all you’ve written is gone. Annoying and difficult to rewrite some good thoughts that I had
<u>App-pos</u>: It helped me see that my negative thoughts are not always high in credibility
<u>App-neg</u>: More personalization</td></tr>
<tr><td>148 (video)</td><td>20</td><td>23</td><td>3</td><td>29</td><td>-</td><td>29</td><td>Low engagement, T2 skipped; depression minimally worse</td></tr>
<tr><td colspan="8"><u>Neg-impact</u>: The anniversary of my dad's passing
<u>App-pos</u>: I liked the where are you in 5 years it was really eye opening
<u>App-neg:</u> Sometimes when it was supposed to mention what I previously wrote it did not work. I think more examples could have been given regarding what could give purpose and so on</td></tr>
</table>

Note. [1]Neg-impact = negative impact on participants’ performance during the intervention (e.g. negative life events); [2]App-pos = positive aspects mentioned about the app; [3]App-neg = negative aspects mentioned about the app; [4]Bug-explanation = technical issues mentioned and explained by the participants that might have impeded their performance.

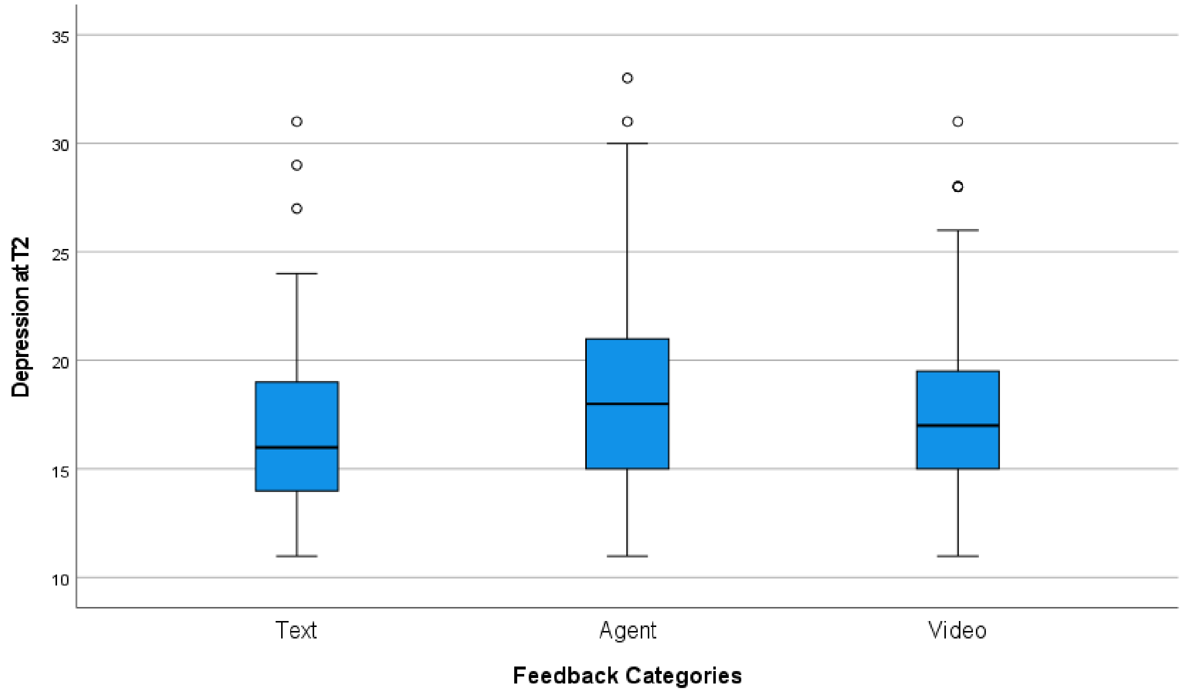
Appendix 4a

Boxplot for depression scores at T1 per feedback category



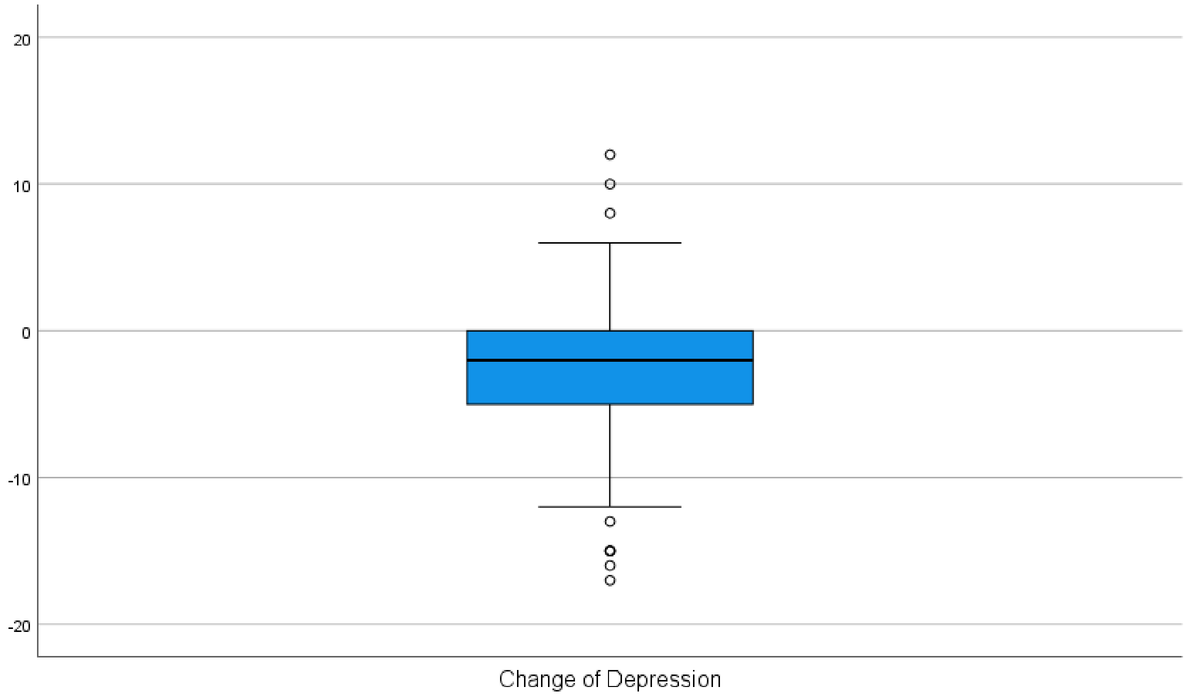
Appendix 4b

Boxplot for depression scores at T2 per feedback category



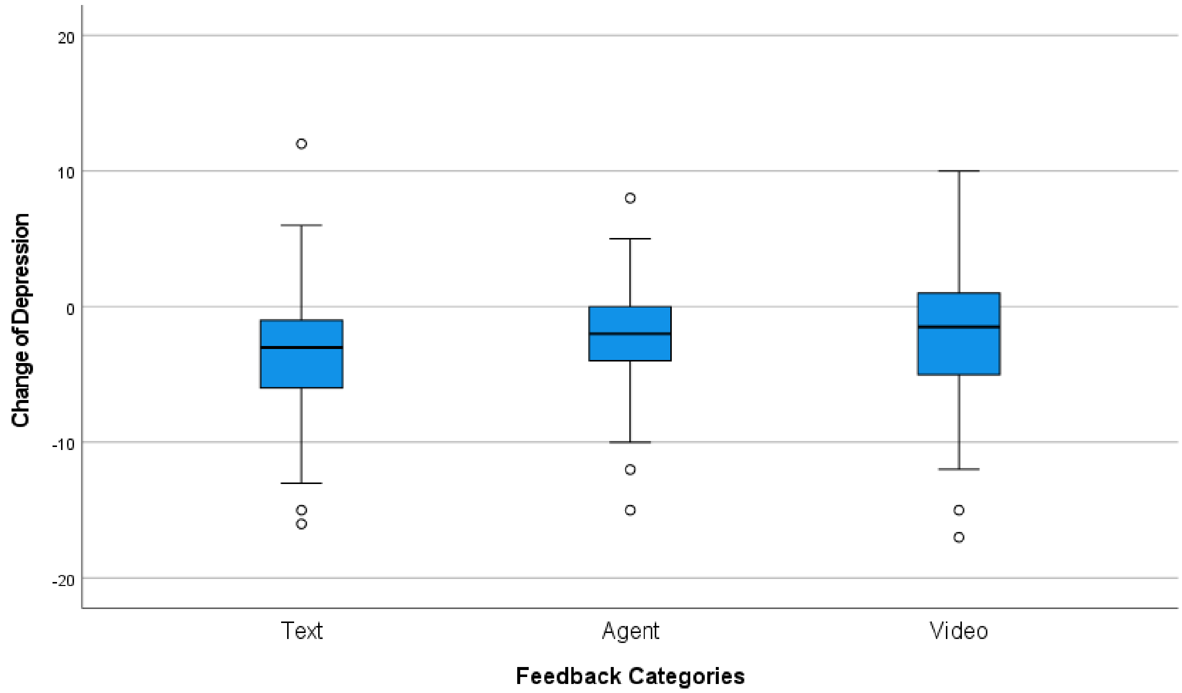
Appendix 4c

Boxplot for the change of depression over time for all participants



Appendix 4d

Boxplot for the change of depression over time per feedback category



Appendix 4e

Boxplot for engagement scores at T1 per feedback category