MASTER

Investigating the impact of personalized treatment goals on engagement levels of SMI patients with an mHealth tool
a pilot study

van Heugten, Jordi M.A.

Award date:
2022

Department of Industrial Engineering and Innovation Sciences

Investigating the impact of personalized treatment goals on engagement levels of SMI patients with an mHealth tool: a pilot study

Master thesis project

by
J.M.A. (Jordi) van Heugten
0912238

Program:	Operations Management and Logistics
First supervisor:	Dr. P.M.E. (Pieter) Van Gorp
Second supervisor:	Dr. R.C.Y. (Raoul) Nuijten
Third supervisor:	Dr. L.E. (Lily) Frank
Fourth supervisor (daily):	L.J. (Lorenzo) James

Eindhoven, July 8, 2022

Investigating the impact of personalized treatment goals on engagement levels of SMI patients with an mHealth tool: a pilot study

Jordi M.A. van Heugten[1], Lorenzo J. James[1], Pieter M.E. Van Gorp[1], Raoul C.Y. Nuijten[1], and Lily E. Frank[1]

ABSTRACT

***Abstract*—Monitoring patients with severe mental illness (SMI) has become a major challenge in mental healthcare. Mobile health (mHealth) tools are more regularly used in a wide range of mental health domains to assess and monitor patients, potentially increasing patient's engagement. Recent results have shown that tailored approaches provide even better results than generic approaches. However, we still lack empirical evidence in the SMI setting. More specifically, it remains unclear how personalized goals, which are critical from a treatment point of view, affect engagement. Therefore, this pilot study aims to evaluate the impact of personalized treatment goals on engagement levels of SMI patients with an mHealth tool. We designed a two-period two-arm within-subject crossover study in which 4 participants were exposed to personalized and non-personalized behavioral goals. It was found that personalized behavioral goals did not have a significant impact on engagement levels, as compared to non-personalized behavioral goals. Additionally, we argued that the goal difficulty might be key when personalizing treatment goals within an mHealth tool. Personalization seems particularly promising when it focuses on dynamically challenging goals for an individual over an extended period of time, balancing the right combination of goals to get a patient into flow.**

***Index Terms*—mHealth, personalization, goal setting, engagement, severe mental illness, FACT.**

I. INTRODUCTION

In recent decades, the number of people who suffer from mental health problems has grown significantly in the Netherlands. Research has shown that more than 40% of the general population will experience one or more mental disorders in their entire lifetime [1]. Individuals who experience mental disorders for an extended period of time (i.e., at least several years) and who have serious limitations in social and societal functioning, are considered to suffer from a severe mental illness (SMI) [2, 3]. It is estimated that approximately 20% of these individuals are at high risk of relapse and hospital readmission [4]. Nevertheless, the majority of the SMI patients are living independently. To prevent social and societal problems, continuous coordination of SMI patients by healthcare professionals is crucial [5]. Therefore, the Dutch Institute of Mental Health and Addiction Care devised Flexible Assertive Community Treatment (FACT), with the aim to treat and support patients with continuity in their own environment in order to decrease admissions and to prevent dropping out of care [2]. During FACT, patients are regularly visited by their case manager (i.e., a healthcare professional), who continuously evaluates and monitors the risk of relapse [2]. These case managers are also responsible for co-designing the treatment outcome goals together with the patients, which are documented in a patient's personal treatment plan [3].

Since the majority of SMI patients are living independently, it has become difficult for case managers to monitor these patients and provide coordinated care. The mental health sector faces a shortage of staff and limited budgets, making it impossible to continuously approach and treat these patients individually at increasing scale [3, 6]. Moreover, due to the recent COVID-19 pandemic, delivery of care at home was not even possible for a while [7]. As a result, patients may not receive the care and treatment they need as there is no effective way for a case manager to monitor the patient [8]. In turn, patients are reportedly disengaged with their treatment, potentially as a result of a poor fit between the patient and their assigned treatment [9]. Even now, there are no effective digital tools to help case managers monitor their patients individually, nor is there a tool to help patients work independently on the outcome goals found in their treatment plans. Current digital tools, such as remote calling or e-mail, were not experienced by many patients as a better alternative, as compared to standard home visits [8]. This indicates that these tools are not suitable for continuous monitoring of treatment goals and motivation to adhere to these goals. Hence, new (digital) approaches for illness self-management, treatment and monitoring are urgently needed [10].

Nowadays, mobile health (mHealth) tools are more regularly used in a wide range of mental health domains, mostly used to tackle lack of engagement, treatment adherence and treatment costs [11, 12]. Previous research has shown some promising results in employing mHealth interventions among SMI patients to positively influence desired behavior change (e.g., compliance with the entire clinical and therapeutic process) [9, 13, 14, 15]. A review of mHealth devices used in clinical interventions concluded that there is an emerging evidence base to support the use of these tools in the assessment, monitoring and intervention of daily functioning of SMI patients [16]. In turn, it improves healthcare delivery, increases diagnosis speed and reduces treatment costs [17].

A behavior change strategy is even more effective when an mHealth tool is personalized toward particular user needs or user characteristics [18, 19]. Empirical studies in non-SMI

[1]Jordi van Heugten, Lorenzo James, Pieter Van Gorp, Raoul Nuijten, and Lily Frank are with Department of Industrial Engineering and Innovation Sciences at Eindhoven University of Technology, The Netherlands (corresponding author: l.j.james@tue.nl)

settings have demonstrated that tailored approaches provide better results than one-size-fits-all approaches, which distribute the same components to all users [20, 21]. Tondello et al. [19] proposed a general framework that can be used to design a system that recommends these components to each user individually. Key in this framework are the three main components that could be tailored [19]: 1) activities, 2) game elements, and 3) persuasive strategies. The absence of empirical results in the SMI setting provides an opportunity to personalize these components within an mHealth tool in order to improve a target behavior. More specifically, the opportunity to personalize the activity component for an individual SMI patient. Since treatment outcome goals are already personalized in consultation with a case manager, mHealth tools could employ these goals to effectively promote a target behavior. Hence, by adopting these personalized goals within an mHealth tool (i.e., personalizing activities), the poor fit between the patient and assigned treatment can be improved, potentially increasing engagement.

Therefore, this study aims to investigate the impact of personalized treatment goals on engagement levels of SMI patients with an mHealth tool. Promoting engagement in mHealth behavior change interventions is thought to be important for intervention effectiveness and is becoming key in mHealth research [9, 19, 22]. Two conceptual models of engagement have been widely used and state that engagement is captured through behavior [23, 24]. In this pilot study, we explored the impact of both personalized and non-personalized tasks (i.e., the activity component) within an mHealth tool on engagement levels. These tasks are comparable to treatment behavioral goals, which, together with treatment outcome goals, were set and documented in a personal treatment plan. The target patients received FACT and are treated by the Dutch Institute of Mental Health and Addiction Care. The treatment related tasks were hand-tailored for each patient by their case manager. Prior studies found that tailored approaches are more effective and provide better results than generic approaches [20, 21]. Therefore, we hypothesized that the impact of receiving personalized treatment goals in an mHealth tool would be larger than the impact of receiving non-personalized treatment goals on engagement levels of SMI patients.

In the remainder of this paper we first survey contemporary mHealth tools for SMI patients and deepen our understanding of behavior change theories. Thereafter, we outline our research methods, which consist of a preliminary desk research and intervention trial including recruitment strategy, intervention context, study design and measurements used for statistical analysis. Subsequently, the obtained results are presented. Finally, we discuss our findings, study limitations, recommend future work directions and draw conclusions.

II. Theoretical Background

Currently, there are no effective digital tools to help case managers monitor their patients and sustain adherence of SMI patients. Therefore, we first survey contemporary mHealth solutions for these patients and how these tools are tailored. Second, we deepen our understanding of behavior change theories. It provides directions on how to improve engagement in settings where case managers co-design treatment goals for mHealth interventions.

A. Contemporary mHealth tools for SMI patients

mHealth tools are increasingly used in a wide range of mental health domains, mostly motivated by a lack of engagement, treatment adherence and treatment costs [11, 12]. To optimally integrate real world elements (e.g., treatment goals), it is essential that these tools are tailored toward a broad range of user needs and user characteristics [25]. Within the concept of mHealth tailoring, a distinction is oftentimes made between customization and personalization. Customization means that a user has the opportunity to adapt the systems content and functionality to their own needs and preferences, while personalization means that a system or person (i.e., not the user itself) offers tailored content or services to a user based on the needs and preferences of that user [11, 18]. Overall, users praised the personalized aspects and customization options, which made the mHealth tool more attractive to them. [26, 27].

As stated before, three main components could be tailored within an mHealth tool. First, several studies suggested that they could benefit from the relationship between the patient and their caregiver and utilize that interaction to tailor activities for the health intervention [28, 29, 30]. Since case managers are responsible for co-designing the treatment outcome goals [3], they should be perfectly able to personalize tasks for such an intervention. This could help to increase the number of tasks conducted by the patient [31]. In turn, case managers should be able to constantly monitor the treatment progress, which may reduce the risk of detrimental effects [32].

Second, the use of gamification in mHealth research has received considerable interest for its potential to increase engagement and target behavior change [11, 14]. Gamification is the application of game elements in non-game environments to promote and affect behavior with gameful experiences [33, 34]. Tailoring these game elements can potentially achieve better results, although this has not yet been supported by empirical evidence [35, 36]. Cheng et al. [11] found that the number of applied gamification elements is growing and that the most used elements in mHealth tools for SMI patients are: levels, narrative or theme, points, rewards and avatars. However, no significant effect of the optimal number of elements on a target behavior was found, while minimal elements (i.e., only 1) were insufficient to promote engagement [37, 38]. A combination of several specific elements seems promising. For example, combining levels with points was found to be associated with favorable feedback by users and increased app usage as it makes the accumulated points more meaningful [25, 39, 40]. On the other hand, a narrative (i.e., virtual therapist) was implemented to improve connectedness between task achievement and progress, which could be interesting because of the important role of a therapist in such a setting [25, 41].

Lastly, several studies applied a tailored persuasive strategy, which are strategies to communicate with a user [19, 30, 42]. For example, based on an individual's current mood state, a personalized empathic message was send. It enabled the mHealth tool to collect mood information in both active (e.g., user logged a mood state) and passive (e.g., smartphone sensor) way and tailor a specific message to it [43].

B. Behavior change theories

With behavior change theories, researchers attempt to explain why human behavior changes. For an mHealth intervention to be effective, several behavior change theories (i.a., COM-B sytem) argue that behavior is a product of three essential conditions: capability, opportunity and motivation [44, 45]. In order to enact a target behavior at a given moment, one must have the capability and opportunity (i.e., environment) to engage in the behavior, and the strength of motivation to engage in the behavior must be greater than for any other competing behaviors [44]. In this 'behavior system', the three components interact to generate behavior that in turn influences these components. For example, capability and opportunity can both influence motivation, which in turn can influence behavior; enacting a behavior can alter capability, motivation, and opportunity [44].

Stimulation of motivation is a key process in behavior change. The COM-B model defines motivation as all those brain processes that energize and direct behavior and distinguishes between automatic processes (e.g., desires and habits) and reflective processes (e.g., plans and evaluations) [44]. An additional framework to study human motivation and personality is the Self-Determination Theory (SDT) [46]. This theory differentiates between intrinsic motivation (i.e., due to internal factors) and extrinsic motivation (i.e., due to external factors) and proposed three basic psychological needs [46]. In mHealth research, a tool enhances intrinsic motivation by satisfying the need of autonomy (i.e., need to feel in control when performing tasks), competence (i.e., need to master tasks and learn different skills), and relatedness (i.e., need to feel connected to others) [46, 47]. In particular, intrinsic motivation drives long-term engagement more than extrinsic motivation [46].

Lastly, specific and challenging goals along with appropriate feedback contribute to higher and better results, according to the Goal Setting Theory [48]. This means that for a person to engage in a target behavior, goals need to be formulated according to five principles: clarity, challenge, commitment, feedback, and task complexity. This emphasizes the need for case managers and patients to jointly define and document acceptable behavioral goals in a patient's personalized treatment plan. Treatment behavioral goals are centered on an individual's action (e.g., going for a walk), while treatment outcome goals focus on a result (e.g., losing 5 kilogram of weight) [49]. Applying these behavioral goals within mHealth interventions could potentially lead to a more successful intervention [50]. By defining specific and challenging goals for an individual (i.e., personalize behavioral goals), the poor fit between the patient and offered treatment can be improved. Thereby potentially increasing engagement, since patients are reportedly disengaged with their treatment due to this mismatch [9].

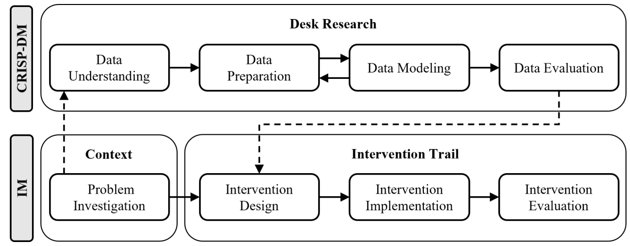


Fig. 1. General research framework, derived from [51] and [52].

III. Methods

General research methodology

To study the impact of personalized treatment behavioral goals within an mHealth tool on engagement levels of SMI patients, this research was conducted in a Dutch mental health context. The Dutch Institute of Mental Health and Addiction Care consists of 15 different FACT teams, each consisting of approximately 8 case managers. The entire FACT population comprised roughly 3,000 SMI patients, distributed over these 15 different FACT teams in the Netherlands. Hence, the caseload of one single case manager is about 25 patients. This also means that all patients of a FACT team have their own case manager. These case managers are responsible for co-designing the treatment outcome goals together with the patients, which are documented in a patient's personal treatment plan.

For this pilot study, we have derived a research framework from two existing frameworks, which is visualized in figure 1. The first framework, the Intervention Mapping (IM) [51] protocol, describes a step-by-step decision-making process for the development, implementation, and evaluation of a health promotion program. After our problem investigation in the Dutch mental health context, we designed our intervention, implemented it in this context, and evaluated the results. During this *intervention trial*, SMI patients were exposed to our designed mHealth application. Since case managers are responsible for co-designing the treatment outcome goals [3], they should be perfectly able to personalize tasks for such an intervention. Therefore, we aimed to design our application with these case managers. During a preliminary *desk research*, secondary data from treatment plans of the Dutch Institute of Mental Health and Addiction Care was analyzed with the objective to recruit case managers who already formulated specific and challenging treatment behavioral goals. Behavioral goals are concrete (i.e., centered on an individual's action [49]), and therefore, measurable. These kind of goals were needed for our intervention design because personalizing the treatment goals within an mHealth tool potentially improves a target behavior of a patient. The second framework, the Cross-Industry standard process for data mining (CRISP-DM) [52] approach (see figure 1), was used to transform the data into useful and actionable information. To extract relevant information for the intervention trial, a text mining method was used that analyzed the current state of the treatment plans for each FACT team. This is particularly useful because people communicate information with language (e.g., texts in treatment plans), which is typically unstructured data. Through text mining, new, previously unknown information is discovered by a computer, which automatically extracts this information from different written resources [53]. Based on our data evaluation of the treatment plans, case managers from a FACT team in Soest/Baarn were approached to participate in the project and to personalize tasks that were needed for our intervention design.

Desk research

A. Data understanding

To personalize the behavioral goals for an mHealth tool, case managers had to be recruited, who already formulated specific and challenging goals. Therefore, secondary data from treatment plans was analyzed to evaluate the current state of these treatment plans and their goals for each FACT team. The data consisted of semi-structured data in a database, including tags that relate to: 1) patient, 2) FACT team, 3) main goal category, and 4) description of main goal. However, the Dutch text within the description of the goals was open-ended and did not have a clear structure. Therefore, text mining and texts analysis were performed using *Python* programming software [54] to obtain quantitative insights related to case managers goal setting strategies.

An exploration of the treatment data was conducted, including reading several treatment plans to understand the data and manually look for patterns. Additionally, a meeting with 3 healthcare professionals was scheduled to clarify each aspect of the treatment plan. Since text mining methods are mostly based on statistical measures of words, a descriptive statistical analysis of the content of the treatment plans was performed. Note that texts, and their statistics, were distorted by text variation. Therefore, texts of treatment plans were preprocessed for further analysis.

B. Data preparation

Before extracting relevant information from the treatment plans, texts (i.e., descriptions of main goals) were preprocessed with different techniques. Words can have different meanings depending on the their context or be read in different ways. First, tokenization was applied, which is a process of splitting a text into words called *tokens*. These tokens were used for the descriptive statistical analysis as well as for further data preparation. Second, the tokens were normalized, including a conversion of all text to lowercase. Third, non-informational text was filtered out, including special characters, punctuation and a list of Dutch most commonly used function words, stopwords and verbs [55]. Eventually, this resulted into a *bag-of-words* for data modelling.

C. Data modelling

1) N-grams: It was assumed that words or sequences of words that occur frequently indicate important content. Therefore, *N-gram* models were used to predict the occurrence of a sequence of *N* words. Since verbs were considered important for behavioral goals, unigrams (i.e., single word) were modelled both with and without inclusion of the list of most commonly used Dutch words. Both lists of unigrams were evaluated by two authors (i.e., J.v.H, and L.J) on the 50 most frequently used words in the description of the treatment plans. Based on the most frequently used unigrams, several words (i.e., *"action"*, *"goal"* or verbs) or sequences (i.e., first-person narratives) were assumed to be important. Thereafter, bigrams (i.e., two words) and trigrams (i.e., three words) were modelled to validate the importance of these words and sequences. Note that data preparation and data modeling are iterative processes and we only modeled bigrams and trigrams including verbs after assuming they were important.

2) Text representation: Two different subset bag-of-words, which were assumed to indicate important content, were used to evaluate FACT teams on their goal setting. While manually evaluating the treatment plans, it was observed that several special characters were used as abbreviation for different important words (e.g., *"a/"* or *"*"* stands for *"action"*). Therefore, the first subset of words included the words *"action"*, *"goal"* and both their observed abbreviations. The second subset of words included sequences of first-person narratives found with N-gram data modelling. These both bag-of-words were used for evaluation.

D. Data evaluation

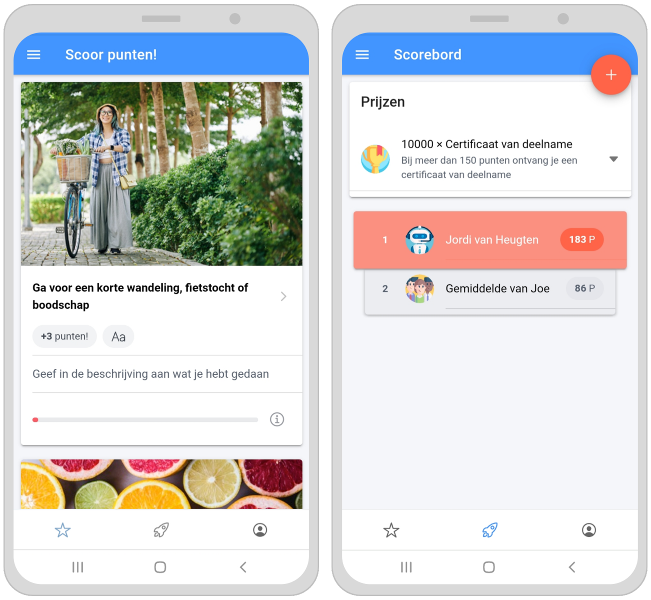
The frequencies of important words and sequences (i.e., the two subset bag-of-words) were evaluated for each FACT team. These frequencies were used as performance indicator, where a higher relative frequency means a higher performance. In other words, a higher performance indicate that the case managers of that specific FACT team were assumed to be more familiar with behavioral goal setting. This study aimed to recruit those case managers who already formulated the most concrete behavioral goals necessary for the design of our mHealth tool. Behavioral goals are measurable and therefore suitable for mHealth interventions. Hence, a FACT team with a higher performance was approached and their case managers were asked to participate in this study.

Intervention Trial

A. Recruitment

1) Case managers: For the design process of the intervention trial, it was essential to first recruit the case managers before actually recruiting the patients. Based on the results of the desk research, case managers from a FACT team in Soest/Baarn, the Netherlands were recruited in March and April 2022.

2) Participants: Participants were recruited among SMI patients who receive treatment from a FACT team in Soest/Baarn, the Netherlands in April and May 2022. Case managers approached the SMI patients they deemed fit and recruited the patients who were willing to participate in the intervention. Thereafter, explicit consent of all participants was collected upon registration for the mHealth program by the researcher, as there may be pressure to consent to participation when consent is collected by the case manager itself. All procedures were also approved by the ethical committee of



(a) (b)

Fig. 2. Interface design of: (a) displayed tasks and (b) the leaderbord with *Average Joe*.

the Eindhoven University of Technology (experiment ID: RB2022IEIS8) and considered as not subjected to Medical Research Involving Human Subjects Act (WMO) by METC Isala Zwolle (experiment ID: 220401 SGP).

B. Intervention context

To evaluate personalized treatment behavioral goals in a Dutch severe mental health context, we have designed our intervention with the mHealth platform *GameBus* (e.g., see www.gamebus.eu). GameBus is a gamification engine that rewards players for playing healthy social, cognitive and physical activities together in a personalized gaming experience. The key idea is to let people pursuit of different health tasks they truly enjoy as an individual. The platform generates integrated health data, in a manner compliant to European privacy legislation, which can be used for scientific research and released its first version late 2015 [56]. Meanwhile, GameBus has been used in a variety of studies related to improving health, such as a similar study which also evaluates the effectiveness of personalization and gamification within mHealth applications, but then related to governmental staff [57].

Considering that GameBus is already a flexible mHealth system with the goal of helping people strive for not only better physical health but also improved mental health, GameBus was used as a basis to design the mHealth intervention for this particular research problem. It can be easily adapted to different studies by changing the web version of GameBus with custom components. Hence, a customized web application was designed to promote engagement of SMI patients by rewarding any performed task (i.e., a behavioral goal within an mHealth tool) with points. Proof of a conducted task was based on a given description by the participant.

The designed application was titled *"Samen Gezond met Joe"* (i.e., *"Healthy together with Joe"*). Title words *"with Joe"* were chosen for specific reasons. The overall goal of the intervention from the perspective of the patients was to obtain as many points as possible by performing treatment related tasks, see figure 2a. Although GameBus had the option to compare user performances on a social leaderbord (i.e., sum of points per user), we did not want participants to see each other's progress due to privacy reasons or other negative associations (e.g., a bad feeling because the participant is at the very bottom of the leaderbord). Instead, participants could only track their own performance and compare themselves against the average performance across all participants (i.e., *Average Joe*) on a leaderboard, see figure 2b. Participants had to get the impressions that they were scoring points *"with Joe"*, rather than against him. To stimulate participants to be actively involved during the intervention, a certificate of participation was awarded to participants that obtained at least 150 points, at the end of the campaign. The entire campaign had a duration of 2 weeks and was split into a personalized week and a non-personalized week.

C. Study design

This study was designed as a two-period, two-arm (2x2) crossover design where each participant was randomized to a sequence of treatments administered sequentially (i.e., within-subject) during treatment periods, although the objective remained a comparison of the two treatments [58]. Such a design is commonly used in clinical trials. An additional advantage is that crossover designs require fewer participants than a parallel design because participants serve as their own control group [58]. This was especially useful because of the limited number of available participants. Figure 3 shows a schematic representation of the study design, including sequences P-NP (i.e., study arm 1) and NP-P (i.e., study arm 2).

In the first week, participants were randomly assigned to either personalized tasks or non-personalized tasks. Whereafter, in the second week, participants were assigned to the other treatment group. The tasks for each week were set in collaboration with the case managers of the participating patients. During a workshop session between these case managers and researchers, a distinction was made between how personalized and non-personalized behavioral goals should be defined. These goals should be clear, specific and, for personalization, based on task complexity, as proposed by the Goal Setting Theory [48]. Then, the goals were objectively

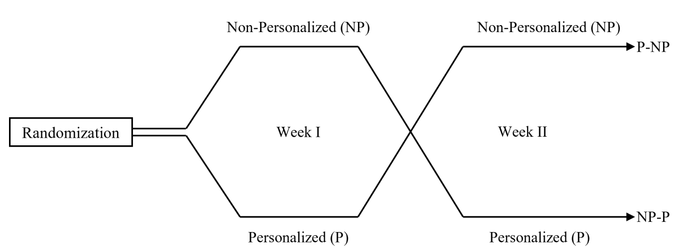


Fig. 3. Study design: 2x2 crossover design.

measurable and suitable as a task within the GameBus application. The workshop presentation slides for the case managers can be retrieved from Figshare [59]. These slides include some guidelines (i.e., examples of personas extracted from the preliminary data analysis) to help case managers formulate behavioral goals for an mHealth tool.

1) Personalized treatment: For each individual patient, a number of personalized behavioral goals were defined by the case manager. Those goals were based either on outcome goals as defined in their personal treatment plan, general lifestyle or social outcome goals. Moreover, these goals were tailored based on task complexity, which implies tailoring the frequency (i.e., how many times performed in a given timeframe) and/or intensity (e.g., for how long, for how far, etc.) [57]. GameBus research suggested that personalization seems particularly promising for promoting the frequency parameter, as opposed to the intensity parameter [57]. Even though those results were preliminary, we lacked better guidelines in international literature.

These personalized behavioral goals were used as personalized tasks within the mHealth tool, with a different number of tasks (T) per patient (i.e., 3 to 6). For each task (t) an importance classification (I) was made by the case manager, which means that tasks were coded between 1 (i.e., least important) and 5 (i.e., most important). Taking into account the frequency (f), participants were able to perform at most between 11 and 35 tasks within the mHealth tool per week. Also the frequency for each task was set by the case manager of the patient. To maintain an equal number of points per participant per week, these parameters were used to calculate the number of points per task (P_t) and the number of points per task taking into account the frequency per week ($P_{t,f}$), see equations 1 and 2. Eventually, participants were able to accumulate roughly 105 points if they would fully comply during the personalized week. Table III, in appendix A displays the lists of the personalized tasks that were suggested to each participant individually.

$$P_t = \frac{I_t}{\sum_{i=1}^{T} I_i} * 105 \tag{1}$$

$$P_{t,f} = round(\frac{P_t}{f_t}) \tag{2}$$

2) Non-personalized treatment: Each case manager defined some non-personalized behavioral goals, from which they thought it was relevant for every patient. Those goals were either lifestyle or socially related. Based on their joint input, a number of non-personalized tasks (i.e., 5 most relevant non-personalized behavioral goals) were selected and approved by the case managers. These tasks were assigned to all participants. As opposed to the personalized tasks, these non-personalized tasks were equally important (i.e., coded with 3) and not tailored according to frequency or intensity. Instead, each task could be performed once a day. Hence, participants were able to perform at most 35 tasks (i.e., accumulate a maximum of 105 points) during the non-personalized week. Table IV, in appendix B displays the lists of the non-personalized tasks that were suggested to all participants.

D. Study procedures

1) Case Managers: Prior to the intervention trial, FACT team Soest/Baarn was approached to participate in the project. A meeting with their team leader and 2 case managers was scheduled to explain about the project and to find out if certain case managers were interested in participating. Thereafter, a similar meeting was scheduled with the 4 case managers who were willing to be part of the project. Up to the kick-off meeting with the actual participants (i.e., patients), we communicated with the case managers via email and phone about patient recruitment and the development of the (non-) personalized behavioral goals and tasks for the mHealth application.

2) Patients: Throughout the intervention period, several emails have been sent and meetings have been scheduled with the participants. At the start of the campaign, a kick-off workshop was scheduled to inform every participant on how to get started with the application. These workshop presentation slides can be retrieved from Figshare [59]. Since the accounts were preconfigured, every participant received an email with personal credentials for the application. Besides, it was requested in the same email to complete the first (pre-test) survey. Additionally, at the end of each week another email have been sent with the request to complete the intermediate-test and post-test survey, respectively. Finally, after the two-week campaign, interviews were scheduled with several participants to evaluate the mHealth application.

E. Measurements

Engagement in mHealth behavior change interventions is thought to be important for intervention effectiveness [9, 22]. Higher patient engagement levels are accompanied with higher patient motivation to adhere to their treatment plan of mental health disorders [15]. In mHealth intervention research, two conceptual models of engagement have been widely used and state that engagement is captured through behavior [23, 24]. Short et al. [22] provided a comprehensive overview on measuring engagement in these interventions, including qualitative measures, self-report questionnaires, system usage data, sensor data, social media data, ecological momentary assessments, and psycho-physiological measures. For this study, measures were collected from our mHealth application (i.e., system usage data), pre-test, intermediate-test, and post-test surveys (i.e., self-report questionnaires), and interviews (i.e., qualitative measures). These measures were often used (in combination) to assess engagement [22]

1) Objective system data: To objectively measure participant engagement, system data was recorded. Yardley et al. [23] distinguishes between micro- and macrolevel engagement when examining the relationship between user experience, usage and behavior change. Microlevel engagement refers to the moment-to-moment engagement with the application, while macrolevel engagement refers to engagement and identification with the wider intervention goals (e.g., actual health behaviors) [23]. Within microlevel engagement, a distinction is made between passive engagement (e.g., visiting the application) and active engagement (e.g., registering a healthy task in the application) [24]. In this study, only microlevel engagement was captured through three variables: 1) the number of days a participant had been online (i.e., passive engagement), 2) the number of tasks a participant had performed (i.e., active engagement), and 3) the number of virtual points a participant had scored (i.e., active engagement), which is considered as a relative scale of task attainment in a particular week. This third variable was introduced to compare individual patients in terms of active engagement, since each participant was assigned to a different number of tasks. For each participant all three measures were recorded per week. Since each week for each patient was related to either personalized or non-personalized tasks, we were able to evaluate the impact of personalized treatment goals on levels of active and passive engagement.

2) Subjective survey data: Three surveys were used to collect subjective data of participants. A pre-test survey was used to gather: 1) demographic data, 2) data related to (intrinsic) motivation, and 3) data related to personality traits. First, this survey recorded participants' age group and gender. Second, intrinsic motivation related to the mHealth application was measured using 4 sub-scales from the Intrinsic Motivation Inventory (IMI) [60]: 1) interest/enjoyment (6 items), 2) perceived choice (7 items), 3) perceived competence (5 items), and 4) pressure/tension (5 items). This multidimensional scale assesses participants' subjective experience related to a target behavior [60]. The interest/enjoyment scale is considered a self-report measure of intrinsic motivation [60]. The perceived choice and perceived competence scales are theorized to be positive predictors of both self-report and behavioral measures of intrinsic motivation [60]. The pressure/tension scale is theorized to be a negative predictor of intrinsic motivation [60]. All items of these 4 sub-scales were measured on 5-point Likert scales (i.e., coded between -2 and +2). Third, Big-5 personality traits were estimated using the mini-IPIP scales [61], which measures a persons' level of 1) openness to experience, 2) conscientiousness, 3) extraversion, 4) agreeableness, and 5) neuroticism with 4 items per trait on 5-point Likert scales (i.e., coded between -2 and +2). This is one of the most commonly used scales in behavioral science to measure personality characteristics. An overview of the pre-test survey questions can be found in appendix C, table V.

Similar surveys (i.e., intermediate- and post-test) were used to measure intrinsic motivation after a participant had receive personalized or non-personalized treatment during one week. Demographic variables and Big-5 personality traits were not measured again. For these surveys, sub-scales of interest/enjoyment and perceived competence consisted of 7 and 6 items, respectively. Both again measured on 5-point Likert scales (i.e., coded between -2 and +2). An overview of the intermediate-test and post-test survey questions can be found in appendix D, table VI.

3) Subjective interview data: At the end of the intervention, semi-structured interviews were conducted to further analyze the motivation of the participant to either be involved in the project or not. All participants were invited to an individual 30-minute interview. The interviews were conducted on location and were supervised by two authors (i.e., J.v.H and L.J.). Unfortunately, due to the participant's personal circumstances, one interview was conducted by filling in the questions via a digital questionnaire. The interview questions focused on: 1) mHealth platform in general, 2) patients' preferences for either the personalized tasks or non-personalized tasks, and 3) motivation. An overview of the interview questions can be found in appendix E, table VII.

F. Data analysis

To evaluate the impact of personalized treatment goals on engagement levels of participants, four different analyses were performed. The statistical analyses were executed using *R* programming software [62]. Statistical tests were two-tailed and a p-value of 0.05 was considered statistical significant. First, an exploration of user statistics was conducted, including descriptive statistics of demographics (i.e., gender, age, personality traits). Additionally, details about the number of participants enrolled in different study phases were provided. Second, statistical analyses were performed to evaluate the impact of personalized treatment goals on engagement levels of participants. These analyses focused on: 1) evaluating passive engagement levels and 2) evaluating active engagement levels. Exploratory analyses were performed, including mean plots and paired samples t-tests to examine potential differences between treatment groups, study arms and age groups. Third, statistical analyses were performed to evaluate the impact of personalized goals on levels of intrinsic motivation, including: 1) interest/enjoyment, 2) perceived choice, 3) perceived competence, and 4) pressure/tension. Again, exploratory analyses were performed using mean plots and repeated measures ANOVA tests, including A Tukey multiple pairwise-comparison, to examine potential differences between treatment groups, including a pre-test condition (i.e., control group). Finally, based on digital recordings of the interviews, interview data was transcribed and organized per question. For each topic (i.e., GameBus web application, goals and motivation) a set of actual quotes was selected that all participants did agree on. These quotes were selected by the first author (i.e., J.v.H). The second author (i.e., L.J) checked these selected citations and agreed with the results.

IV. Results

Desk research

A. Descriptive statistics

Secondary data was retrieved from the Dutch Institute of Mental Health and Addiction Care in February 2022 and included 14,128 data points of treatment plans from January 2010 until February 2022. During that period, 3,392 different patients were treated by 15 different FACT teams, each with a treatment plan containing on average 4.17 (main) long-term treatment goals. These goals were divided into 7 categories: 1) general or psychotherapy (28.0%), 2) social (21.1%), 3) health recovery and symptoms (16.5%), 4) personal lifestyle

(12.3%), 5) functional and daily skills (10.9%), 6) medical or paramedical (9.2%), and 7) diagnostic research (2.0%). Note that the non-personalized goals were either lifestyle or socially related, and therefore, correspond to some of treatment outcome goals documented in a patient's treatment plan. For each long-term goal (i.e., data point) a description was added in the treatment plan. Since this was mostly textual data, tokenization of all descriptions resulted into 384,569 tokens, including 27,142 unique words. Hence, a token ratio of 0.071, which implies that the descriptions of the long-term goals were relatively simple, rather than specific. Additionally, the average length (i.e., number of tokens) of these descriptions differed per FACT team (i.e., $\mu = 25.4, \sigma = 8.1$).

B. Evaluation of treatment plans

It was observed that words forming the first-person narrative (e.g., *"ik ga"*, which translates to *"I go"*) and the words *"actie(s)"* (i.e., action(s)) and *"doel"* (i.e., goal) were often transcribed in the treatment plan (see table I for the most relevant unigrams after tokenization). Additionally, it was observed that several special characters were used as abbreviation for different important words (e.g., *"a/"* or *"*"* stands for *"action"*). The modeled bigrams and trigrams, validated the importance of the first-person narratives (see table VIII, appendix F for most relevant bigrams and trigrams after tokenization). Note that in both cases the most commonly used Dutch words were included.

Figure 4 shows the percentages (i.e., relative frequencies) of the treatment plans that contain at least one specific N-gram for each of the bag-of-words. Results showed that especially the frequency of the first bag-of-words was different for each FACT team. On average, 22.5% of the descriptions contained at least one token of that set of words and only 4 of the FACT teams were performing better than average. This may indicate that the treatment plans often did not contain a clear structure for a specific and challenging behavioral goal (e.g., *"goal: I am going for a walk"*), and therefore, are not directly suitable as input for the intervention design. The frequency of the second bag-of-words was more equally distributed among the FACT teams. On average, 56.4% of the descriptions contained at least one of the first-person narratives from that set of sequences. However, it was not that clear whether all these

TABLE I
UNIGRAMS OF TREATMENT PLANS

	Including [1]			Excluding [1]	
Index	Unigram	Occurrence	Index	Unigram	Occurrence
1	ik	23,981	3	ga	1,527
9	wil	5,718	5	actie	1,206
13	heb	3,658	12	werk	1,047
18	kan	2,528	26	acties	688
32	ga	1,527	36	doel	571
33	ben	1,427			
39	actie	1,206			
46	werk	1,047			

[1] Most commonly used Dutch words.

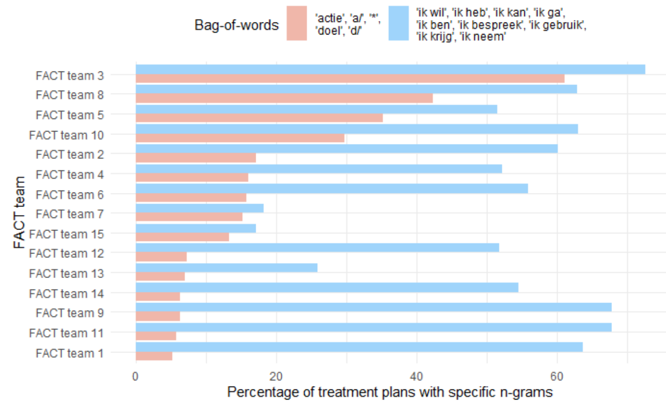


Fig. 4. Percentages of treatment plans that contain a specific bag-of-words.

first-person narratives were related to a treatment behavioral goal or not. For example, when having a closer look at the treatment plans, it was observed that a first-person narrative might also be related to a treatment outcome goal (e.g., *"I would like treatment for my addiction"*). Overall, these results suggest that most case managers did not document specific and challenging behavioral goals in a patient's treatment plan, which were necessary for intervention design. Nevertheless, FACT team 5 (i.e., Soest/Baarn), which scored better than average based on the frequencies of the first bag-of-words, was approached for participating in the intervention trial. It was assumed that case managers from this particular FACT team could more easily design behavioral goals for the mHealth intervention, compared to case managers from other FACT teams which performed below average.

Intervention Trial

A. User statistics

In total, 5 unique participants were enrolled in this study. One participant decided not to give informed consent for the collection and application of system usage data and surveys.

TABLE II
SAMPLE DEMOGRAPHICS

Characteristic	Sample (N = 4)
Gender (N; %)	
Female	3; 75%
Male	1; 25%
Age group (N; %)	
35-44	2; 50%
45-54	2; 50%
Personality ($\mu; \sigma$)	
Extraversion	-0.188; 0.875
Agreeableness	1.250; 0.456
Conscientiousness	0.625; 0.629
Neuroticism	0.313; 1.125
Openness	0.563; 0.239

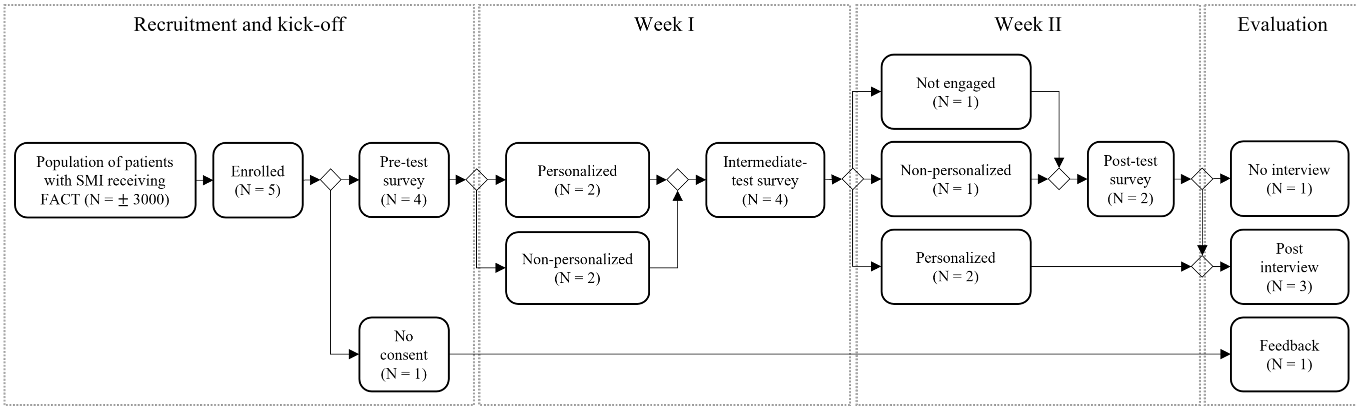


Fig. 5. Cohort diagram.

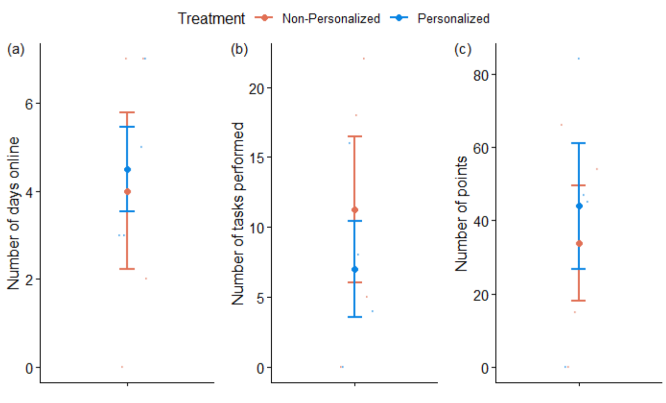


Fig. 6. Mean plots of: (a) the number of days participants had been online, (b) the number of tasks participants had performed, and (c) the number of points participants had scored.

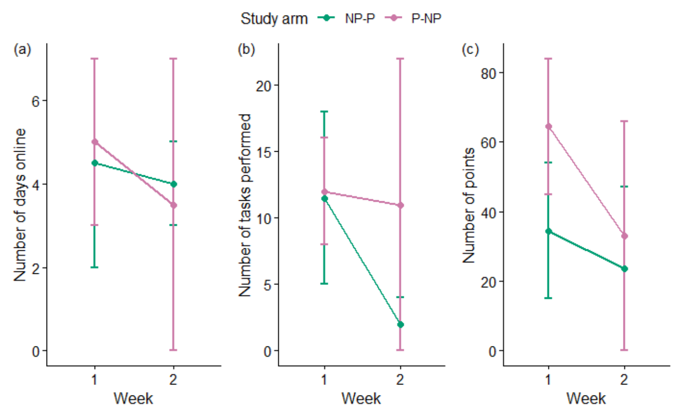


Fig. 7. Mean plots of: (a) the number of days participants had been online, (b) the number of tasks participants had performed, and (c) the number of points participants had scored.

Instead, some feedback was given, including consent to only use this feedback. The remaining participants were randomly assigned to either personalized tasks or non-personalized tasks, in the first week. These 4 participants completed the pre-test survey, performed at least one task during the first week, and completed the intermediate-test survey. During the second week, 1 participant who was assigned to non-personalized tasks, was not engaged (i.e., did not check the application and did not perform a task). At the end of that week, both participants with non-personalized tasks completed the post-test survey, while both participants with personalized tasks did not. At the end of the campaign, 3 participants completed the post-interview. Note that, due to a participant's personal circumstances, one interview was conducted by filling in the questions via a digital questionnaire. Figure 5 shows a cohort diagram which details the number of participants involved in each study phase.

Sample demographics based on the responses of the pre-test survey are displayed in table II. Both age groups were equally distributed in both study arms. Note that these results are based on the 4 participants that gave explicit written consent.

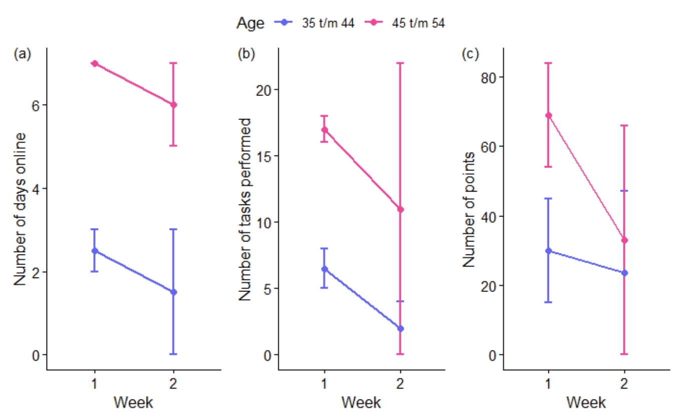


Fig. 8. Mean plots of: (a) the number of days participants had been online, (b) the number of tasks participants had performed, and (c) the number of points participants had scored.

B. Analysis of objective measures of engagement

1) Evaluation of passive engagement levels: Figure 6a suggests no significant differences in passive engagement levels between both treatment groups. However, it seems that there is more variation between participants during the non-personalized week. A paired samples t-test revealed that treatment groups were indeed not statistically different from each other in terms of passive engagement levels. Figure 7a displays passive engagement levels, per week, per study arm. No statistical differences were found between both study arms, while passive engagement seems to decrease over time in general. Visual inspection suggests that the average number of days online decreased faster over time when a participant changed from personalized tasks to non-personalized tasks (i.e., P-NP), as compared to a change from non-personalized tasks to personalized tasks (i.e., NP-P). Nevertheless, these results were also not significant. Lastly, figure 8a displays passive engagement levels, per week, per age group. A visual inspection suggests that there are differences between both age groups in terms of passive engagement, but t-test showed no significant differences in both the first (i.e., $p = 0.070$) and second (i.e., $p = 0.148$) week.

2) Evaluation of active engagement levels: Figures 6b and 6c suggest no significant differences in active engagement levels between both treatment groups, which is confirmed with paired samples t-tests. Figure 7b and 7c display active engagement levels, per week, per study arm. Again, no statistical differences were found between both study arms. Also active engagement seems to decrease over time in general. The same decreasing effect over time as with passive engagement was observed for the number of points obtained, per study arm, although again not significant. Regarding the number of tasks performed it was observed that this decreased faster over time when a participant changed from non-personalized tasks to personalized tasks (i.e., NP-P). This might be explained by the fact that the number of personalized tasks were different for each participant, and therefore this result may be biased. Nevertheless, also this result was not significant. Lastly, figure 8b and 8c display active engagement levels, per week, per age group. Only for the first week, a statistical significant difference was found for the number of tasks performed (i.e., participants in age group 45-54 performed 10.5 tasks more at $p = 0.038$), while no significant difference was found for the number of points obtained ($p = 0.207$). Again, this result may be biased due to an inconsistent number of personalized tasks per participant. Based on this evaluation of active engagement levels, we conclude that only age in the first week leaded to significant differences in terms of number of tasks performed.

C. Analysis of subjective measures of engagement

1) Evaluation of survey data: Figure 9 displays mean plots of the 4 dimensions of intrinsic motivation per treatment group. Before the intervention, responses of a pre-test survey were collected and served as a control group. Visual inspection suggests that participants enjoyed the application less and experienced more tension, after they received personalized treatment. One-way repeated measures ANOVA tests revealed that treatment groups were significantly different from each other in terms of tension ($p = 0.022$). A Tukey multiple pairwise-comparison revealed that personalized treatment was found to have a significant higher level of tension compared to non-personalized treatment (i.e., 0.70 higher at $p = 0.023$) and the control group (i.e., 0.65 higher at $p = 0.032$).

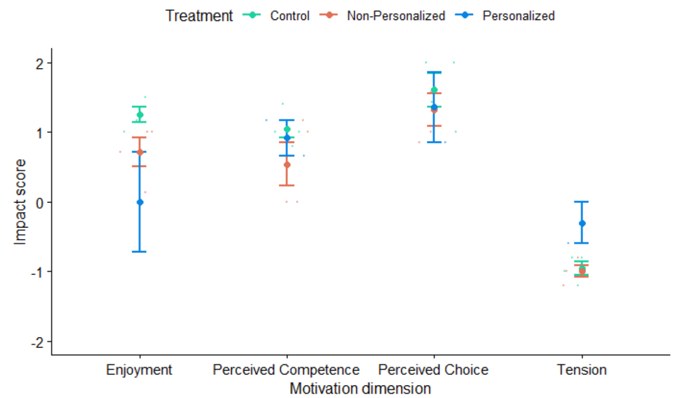


Fig. 9. Mean plots of the impact scores of the dimensions of motivation: 1) enjoyment, 2) perceived competence, 3) perceived choice and, 4) tension.

2) Evaluation of interview data: Of all enrolled participants, three (i.e., 1, 2 & 4) participants completed the post-interview, including one (i.e., participant 4) via a questionnaire due to personal circumstances. On average, they awarded the current version of the GameBus web application an 8 out of 10. The interview has not been conducted with participant 4, who did not respond to the invitation. The fifth participant, who did not give explicit consent to participate, did provide some feedback which we were allowed to use for the study.

Participants agreed that they *"did enjoy the GameBus application"*. Their main reason to use the application was *"to achieve their personal goals"*. Participants 1 & 4 mentioned: *"GameBus helps you with that little bit of extra motivation to get started with your goals"*. Additionally, participant 2 mentioned: *"I was supported by a visual overview of my personal goals"*. In general, the addition of points was experienced as fun: *"I really liked the points, it stimulated me to use GameBus every day"*. However, *"I would like to level up after collecting several points, then the goals may also become more challenging as I progress"*. Participant 4 mentioned that *"the point rating was just the same for each goal"* (i.e., non-personalized). Participant 5 added: *"all you get is a few points against a non-player character in a game that is theoretically possible to cheat"*. As for the current version of the application, *"the basics were just right"* and *"the application was immediately ready for use"*. Participant 1 mentioned that he *"would like to have social contact with both healthcare experts and other patients"*. He added: *"In the form of a public and closed news feed, so you can decide for yourself what you do or do not want to share"*. Both participant 1 & 2 mentioned that the application could be improved by *"implementing avatars with a clear storyline"*. For example: *"First you see an avatar with a really big belly, and after you achieve a goal, he gets a six pack"*. Unfortunately, not all participants were able to use the application on a daily basis. For example, due to health related circumstances or *"not having the consistency or focus to do this every day"*. Nevertheless, participant 1 concluded with: *"More challenge and more features, then it becomes interesting and it remains more interesting"*.

All participants that joined the intervention campaign had a clear preference for the personalized tasks. Participant 1 mentioned that he *"liked the personalized tasks more, these are my own things after all"*. He added that *"it's all about finding ways that suit me and what works best for me. These tasks worked quite well, it could probably work out well in other areas of life"*. However, he was *"unable to complete all personalized tasks. I usually want to do everything right, which makes it difficult then"*. Participant 2 was generally very satisfied: *"Although both sets of tasks were well put together, I prefer the personalized tasks, because these were more applicable to me. In addition, these tasks come naturally"*. Also participant 4 had a preference for the personalized tasks *"because that is were my interests lie"*. Nevertheless she did not perform those tasks due to personal circumstances. Participant 5 added: *"It only works for me if I can really do something with it on my own. It must mean something personal"*.

The non-personalized tasks were generally not challenging enough: *"I did not find these tasks challenging enough. For example, you already brush your teeth, eat healthy meals and fruit every day"*. Participant 1 mentioned that *"these tasks were already quite present in my daily structure"*. Participant 4 added: *"This week everything was more or less the same. More variety will be more fun"*. Nevertheless, participants enjoyed some of these tasks as well. For example, participant 2 mentioned that *"more general lifestyle tasks are also important. Simple things that need to be emphasized again"*.

Overall, participants were satisfied with the use of the application and indicated that they *"had done well"*. *"It has helped me a lot, also with regard to the progress of my treatment. With GameBus, it is easier to start with a goal and stick to it on a regular basis"*. *"It encourages you to complete some challenging tasks"*. Participant 1 mentioned that *"personalized, challenging tasks and points, in combination with levels and avatars, are likely to be motivating. Besides, I think you can really make something beautiful out of it and help a lot of people"*. Two participants explicitly mentioned that they would like to continue using the GameBus web application.

V. Discussion

A. Principal findings

In this pilot study we evaluated the impact of personalized treatment goals on engagement levels of SMI patients with an mHealth tool. From our exploratory statistical analyses, we found that personalized tasks did not have a significant impact on both passive and active engagement levels, as compared to non-personalized tasks. Additionally, engagement levels with the mHealth tool tended to decrease over time for both study arms, although these results were also not significant. It was observed that both passive (i.e., number of days online) and active (i.e., only the number of points obtained) engagement decreased faster over time when a participant changed from personalized tasks to non-personalized tasks, as compared to a change from non-personalized tasks to personalized tasks. This implies that, an individual tends to be more engaged with the mHealth tool when receiving personalized tasks. This reflects with the findings of the post-interviews, in which participants unanimously expressed a clear preference for the personalized tasks. Hence, our hypothesis that the impact of receiving personalized treatment goals in an mHealth tool would be larger than the impact of receiving non-personalized treatment goals was partially accepted.

One interesting finding is that the level of difficulty (i.e., challenge) might be key when personalizing treatment goals within an mHealth tool. Surprisingly, we found that participants rated interest or enjoyment lowest and pressure or tension highest after they received personalized treatment. This implies that patients were least intrinsically motivated to engage with the mHealth tool since both dimensions are a self-report measure and negative predictor for intrinsic motivation, respectively [60]. Only significant differences in levels of tension were observed after a patient received personalized tasks, as compared to receiving non-personalized tasks or the control condition. A potential explanation for this might be that a personalized task may be too challenging to complete, which one participant also mentioned during the post-interview. As a result, someone might get tense because he is not able to complete all tasks despite the desire to do so. When having a closer look at the personalized tasks, someone might argue that these tasks are indeed too specific (e.g, *"Eat 4 sandwiches tonight instead of 6 sandwiches"*. This indicates that the behavioral goals that were set by a case manager were probably too difficult to complete, which potentially harms engagement. Moreover, participants indicated that the non-personalized tasks were generally not challenging enough. Since all tasks were static and not dynamic, the difficulty of a task did not increase or decrease. This implies that tasks were either too difficult (i.e., personalized tasks) or too easy (i.e., non-personalized tasks) to complete. According to Flow Theory [63], the trade-off between challenge and skills must be in balance for a person to be in flow. It was even mentioned that tasks should become more challenging as a person progresses, possibly in the form of a level system. However, in this static version of the GameBus web application, the tasks were not continuously adapted based on this trade-off, and therefore, participants were not continuously in flow. Hence, for an individual to be more engaged with an mHealth tool, tasks should be updated continuously according to skills of that individual. According to the Self-Determination Theory (SDT) [46], an individual is more likely to achieve their goals if the skills (i.e., competence) are perceived as appropriate. By satisfying the need to master tasks and learn different skills, it yields improved self-motivation and mental health [46].

Furthermore, some other aspects may have influenced the engagement levels of participants. From our exploratory statistical analysis, we found that the engagement levels of different age groups were slightly different. It was observed that the oldest age groups were more passive and active engaged, although this was generally not significant and there were only two different age groups. Nevertheless, a significant difference was found in the number of tasks performed, per age group, in the first week. It should be noted that the participants received an inconsistent number of personalized tasks, which may bias these results. This finding was also reported by different mHealth studies for SMI patients, which showed that the age of a person may predict engagement with the application [27, 64]. However, the current finding is contrary to previous studies which have suggested that an mHealth tool is particularly effective in addressing the problem of lack of motivation for youth and young adults [65, 66]. Additionally, some participants mentioned that they were not able to perform certain tasks or stay engaged with the mHealth tool due to their mental illness or other personal circumstances. This implies that the current version of the application does not seem to work for everyone in this target group.

Lastly, results from preliminary desk research suggest that the current goals in a treatment plan are often not behavioral oriented. Treatment plans contain long-term outcome goals, often combined with unstructured description data. This indicates that, case managers have no clear protocol of defining specific and challenging behavioral goals. In other words, each case manager is allowed to define and document behavioral and/or outcome goals in the treatment plan based on their own routine or experience. This is in line with the results of this study, which confirm that the length of these descriptions varies by FACT team and descriptions are relatively simple instead of specific. This means that, (personalized) tasks for an mHealth tool cannot be extracted directly from a treatment plan. Goals which are not behavioral oriented are difficult to measure, and therefore do not fit within an mHealth tool. Nevertheless, several FACT teams used specific words or sequences (i.e., assumed to be important content) in the descriptions of the treatment plans. These teams clearly indicated an action or goal, often followed by the first person narrative. This came closest to setting behavioral goals, and therefore these case managers were approached for participation. This current strategy, including case manager recruitment, may not be the most optimal protocol to select behavioral goals for an mHealth tool. Hence, a more structured way of defining and documenting goals in the treatment plan is desired.

B. Study limitations

This study was subjected to several limitations. First, this study has low power due to a very small sample size (i.e., N = 4), as compared to the entire population which comprised roughly 3,000 SMI patients who receive FACT. Therefore, this study was designed as a crossover design, in which participants served as their own control group. Nevertheless, based on this sample size it was very difficult to find significant results. Additionally, not all participants completed the post-test-survey and post-interview, further reducing the sample size for various analyses. Although participation was voluntary, measures were taken to direct patients to participate in all study phases (e.g., reminders for the surveys).

Second, case managers had to be recruited before patients could eventually be approached. Not all case managers were willing to participate in this study, which automatically resulted in a significant number of patients being excluded. Additionally, the case managers who did participate in this study only approached patients they believed would be interested in using a digital health tool. Hence, our sample also included bias, which may potentially affect the validity of the results. Based on these first two limitations, the sample size could not be increased by the researchers, despite the crossover study design.

Lastly, this pilot study focused on a specific target group in a specific context (i.e., Dutch SMI patients who receive FACT in a healthcare setting). Therefore, the findings in this study could not be generalized to other groups of people or contexts. Especially because of the very small sample size.

C. Future work

Future work should focus on how personalization could be designed over a longer period of time. While our results were generally not significant, they were in line with our expectations for personalizing treatment goals in an mHealth tool. A follow-up study should consider a more dynamic goal setting strategy, in which tasks are continuously updated according to the skills and needs of a patient. A combination of points with level systems could be used to amplify this strategy. Then, the accumulated points are not only more meaningful [25, 39, 40], but it enhances the trade-off between challenge and skill (i.e., Flow theory [63]). Additionally, the application itself could become more dynamic by putting more emphasis on the overall background story (e.g., updated versions of *Joe* as a person progresses). It was mentioned that with challenge and features, such as avatars with a clear storyline, the application may become more interesting. It would even be possible to personalize such a background story, which in turn is closely related to personalizing a persuasive strategy [19].

Unfortunately, case managers were not perfectly able to define personalized tasks with a balanced amount of difficulty. Therefore, future work should focus on automated decision support, in which the difficulty of tasks is continuously updated by an automated system. It may be that a combination of personalized tasks with non-personalized tasks is more appropriate. This could be realized through an opt-in or opt-out mechanism, where patients can choose the tasks themselves (i.e., customization). This study found that participants were generally positive about the application in which they received personalized and non-personalized tasks.

Furthermore, goals in treatment plans are generally set for a longer period of time (e.g., one year). Therefore, the intervention period should be increased in order to evaluate the impact of an mHealth tool on long term engagement levels. To effectively execute and evaluate this, future studies should focus on collecting more data, and therefore an effective strategy to recruit more patients. A study with more enrolled participants may potentially find a significant difference between both treatments in terms of active and passive engagement levels.

Finally, future research should focus on creating a framework or protocol that allows case managers to set treatment behavioral goals in a more structured way, which eventually can be used as a task in an mHealth tool.

VI. Conclusions

This pilot study aimed evaluate the impact of personalized treatment goals on engagement levels of SMI patients with an mHealth tool. It was found that personalized behavioral goals did not have a significant impact on engagement levels, as compared to non-personalized behavioral goals. However, this study also found that tension was rated higher after a participant received personalized treatment, while non-personalized treatment was generally not challenging enough. This implies that the difficulty of a task might be key when a task is personalized within an mHealth tool. Therefore, personalizing treatment goals still seems to have great potential. Future research should focus on dynamically challenging goals for an individual over an extended period of time, balancing the right combination of goals to get a patient into flow.

Acknowledgment

The authors would like to personally thank the following people from the Dutch Institute of Mental Health and Addiction Care for their assistance in bringing this research paper together: Barbara Montagne, Muriel Hagenaars and Harold van Driel. Furthermore, we would like to thank the case managers and patients who participated in this pilot study, without whom this research would have not been possible.

Informed Consent Statement

Informed consent was obtained from all participants that were involved during the intervention trial (i.e. 4). One participant did not give explicit consent for the collection and application of system usage data and surveys, but agreed to use their feedback for evaluation of the application.

Data Availability Statement

Due to privacy statements, the raw data used in this study are not openly available. The workshop slides with both the case managers and the patients can be retrieved from Figshare at https://doi.org/10.6084/m9.figshare.20226135, reference number [59].

Conflicts of Interests

The authors declare no conflict of interest.

Abbreviations

CRISP-DM	Cross-Industry standard process for data mining.
FACT	Flexible Assertive Community Treatment.
IM	Intervention Mapping.
mHealth	mobile health.
SDT	Self-Determination Theory.
SMI	Severe Mental Illness.
WMO	Medical Research Involving Human Subjects Act.

References

[1] R. de Graaf, M. ten Have, and S. van Dorsselaer, *De psychische gezondheid van de Nederlandse bevolking*. 2010.

[2] J. R. Van Veldhuizen, "FACT: A Dutch version of ACT," *Community Mental Health Journal*, vol. 43, no. 4, pp. 421–433, 2007.

[3] GGZ Standaarden, "Generieke module EPA (Ernstige Psychische Aandoeningen)," 2021.

[4] G. Bond, M. Salyers, A. Rollins, and L. Moser, *Toekomstige ontwikkelingen van ACT*. Uitgeverij Cure and Care Publishers Nijmegen, 2005.

[5] P. Delespaul, J. van Weeghel, and C. G. EPA, "Consensus over de definitie van mensen met een ernstige psychische aandoening (EPA) en hun aantal in Nederland," *Tijdschrift voor Psychiatrie*, vol. 55, no. 6, pp. 427–438, 2013.

[6] B. van der Hoek, "Een mentaal gezonde samenleving en de rol van de ggz," tech. rep., Trimbos, 2021.

[7] A. van Giessen, A. de Wit, C. van den Brink, K. Degeling, C. Deuning, J. Eeuwijk, C. van den Ende, I. van Gestel, R. Gijsen, P. van Gils, M. IJzerman, I. de Kok, G.-J. Kommer, L. Kregting, E. Over, A. Rotteveel, K. Schreuder, N. Stadhouders, and A. Suijkerbuijk, "Impact van de eerste COVID-19 golf op de reguliere zorg en gezondheid," *Rivm*, p. 158, 2020.

[8] A. De Lange, L. Hulsbosch, and A. Knispel, "Impact van de coronacrisis op mensen met ernstige psychische aandoeningen," *Utrecht: Trimbos-instituut*, p. 28, 2020.

[9] M. Fitzgerald and G. Ratcliffe, "Serious games, gamification, and serious mental illness: A scoping review," *Psychiatric Services*, vol. 71, no. 2, 2020.

[10] M. Varnfield, K. Rajesh, C. Redd, S. Gibson, L. Gwillim, and S. Polkinghorne, "Health-e Minds: A Participatory Personalised and Gamified mHealth Platform to Support Healthy Living Behaviours for People with Mental Illness," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 6943–6947, 2019.

[11] V. W. S. Cheng, T. Davenport, D. Johnson, K. Vella, and I. B. Hickie, "Gamification in apps and technologies for improving mental health and well-being: Systematic review," *JMIR Mental Health*, vol. 6, no. 6, pp. 1–15, 2019.

[12] S. Carlier, D. Coppens, F. De Backere, and F. De Turck, "Investigating the influence of personalised gamification on mobile survey user experience," *Sustainability (Switzerland)*, vol. 13, no. 18, 2021.

[13] J. A. Naslund, L. A. Marsch, G. J. McHugo, and S. J. Bartels, "Emerging mHealth and eHealth interventions for serious mental illness: A review of the literature," *Journal of Mental Health*, vol. 24, no. 5, pp. 321–332, 2015.

[14] D. Johnson, S. Deterding, K. A. Kuhn, A. Staneva, S. Stoyanov, and L. Hides, "Gamification for health and wellbeing: A systematic review of the literature," *Internet Interventions*, vol. 6, pp. 89–106, 2016.

[15] N. Sinha, *Introducing Gamification for Advancing Current Mental Healthcare and Treatment Practices*, vol. 933. Springer Singapore, 2021.

[16] L. Jameel, L. Valmaggia, G. Barnes, and M. Cella, "mHealth technology to assess, monitor and treat daily functioning difficulties in people with severe mental illness: A systematic review," *Journal of Psychiatric Research*, vol. 145, no. July 2021, pp. 35–49, 2022.

[17] N. El-Rashidy, S. El-Sappagh, S. M. Riazul Islam, H. M. El-Bakry, and S. Abdelrazek, "Mobile health in remote patient monitoring for chronic diseases: Principles, trends, and challenges," *Diagnostics*, vol. 11, no. 4, pp. 1–32, 2021.

[18] R. Orji, L. E. Nacke, and C. Di Marco, "Towards personality-driven persuasive health games and gamified systems," *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-May, pp. 1015–1027, 2017.

[19] G. F. Tondello, R. Orji, and L. E. Nacke, "Recommender systems for personalized gamification," *UMAP 2017 - Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, no. July, pp. 425–430, 2017.

[20] W. Hochleitner, M. Lankes, L. E. Nacke, M. Tscheligi, M. Busch, E. Mattheiss, R. Orji, and A. Marczewski, "Personalization in serious and persuasive games and gamified interactions," *CHI PLAY 2015 - Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, pp. 811–816, 2015.

[21] R. Orji, A. Dijkstra, M. Busch, M. Reisinger, E. Mattheiss, and M. Kaptein, "Personalization in Persuasive Technology," *Adjunct Proceedings of 12th Persuasive Technology conference*, no. April, 2017.

[22] C. E. Short, A. DeSmet, C. Woods, S. L. Williams, C. Maher, A. Middelweerd, A. M. Müller, P. A. Wark, C. Vandelanotte, L. Poppe, M. D. Hingle, and R. Crutzen, "Measuring engagement in eHealth and mHealth behavior change interventions: Viewpoint of methodologies," *Journal of Medical Internet Research*, vol. 20, no. 11, pp. 1–18, 2018.

[23] L. Yardley, B. J. Spring, H. Riper, L. G. Morrison, D. H. Crane, K. Curtis, G. C. Merchant, F. Naughton, and A. Blandford, "Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions," *American Journal of Preventive Medicine*, vol. 51, no. 5, pp. 833–842, 2016.

[24] O. Perski, A. Blandford, R. West, and S. Michie, "Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis," *Translational Behavioral Medicine*, vol. 7, no. 2, pp. 254–267, 2017.

[25] P. Siriaraya, V. Visch, M. M. Van Dooren, and R. Spijkerman, "Learnings and challenges in designing gamifications for mental healthcare: The case study of the readysetgoals application," *2018 10th International Conference on Virtual Worlds and Games for Serious Applications, VS-Games 2018 - Proceedings*, 2018.

[26] J. Milward, P. Deluca, C. Drummond, R. Watson, J. Dunne, and A. Kimergård, "Usability testing of the BRANCH smartphone app designed to reduce harmful drinking in young adults," *JMIR mHealth and uHealth*, vol. 5, no. 8, 2017.

[27] R. Tark, M. Metelitsa, K. Akkermann, K. Saks, S. Mikkel, and K. Haljas, "Usability, acceptability, feasibility, and effectiveness of a gamified mobile health intervention (triumf) for pediatric patients: Qualitative study," *JMIR Serious Games*, vol. 7, no. 3, 2019.

[28] P. Lindgreen, K. Lomborg, and L. Clausen, "Patient experiences using a self-monitoring app in eating disorder treatment: Qualitative study," *JMIR mHealth and uHealth*, vol. 6, no. 6, pp. 1–18, 2018.

[29] M. M. van Dooren, P. Siriaraya, V. Visch, R. Spijkerman, and L. Bijkerk, "Reflections on the design, implementation, and adoption of a gamified eHealth application in youth mental healthcare," *Entertainment Computing*, vol. 31, no. January, p. 100305, 2019.

[30] M. H. Lim, J. F. Gleeson, T. L. Rodebaugh, R. Eres, K. M. Long, K. Casey, J. A. M. Abbott, N. Thomas, and D. L. Penn, "A pilot digital intervention targeting loneliness in young people with psychosis," *Social Psychiatry and Psychiatric Epidemiology*, vol. 55, no. 7, pp. 877–889, 2020.

[31] J. Boettcher, K. Magnusson, A. Marklund, E. Berglund, R. Blomdahl, U. Braun, L. Delin, C. Lundén, K. Sjöblom, D. Sommer, K. von Weber, G. Andersson, and P. Carlbring, "Adding a smartphone app to internet-based self-help for social anxiety: A randomized controlled trial," *Computers in Human Behavior*, vol. 87, pp. 98–108, 2018.

[32] M. Kocur, M. Dechant, C. Wolff, C. Nothdurfter, T. C. Wetter, R. Rupprecht, and Y. Shiban, "Computer-Assisted Avatar-Based Treatment for Dysfunctional Beliefs in Depressive Inpatients: A Pilot Study," *Frontiers in Psychiatry*, vol. 12, no. July, pp. 1–13, 2021.

[33] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining 'Gamification'," *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*, pp. 9–15, 2011.

[34] K. Huotari and J. Hamari, "Defining Gamification - A Service Marketing Perspective," *Proceedings of the IADIS International Conference Interfaces and Human Computer Interaction 2012, IHCI 2012, Proceedings of the IADIS International Conference Game and Entertainment Technologies 2012*, pp. 227–231, 2012.

[35] G. F. Tondello, R. R. Wehbe, L. Diamond, M. Busch, A. Marczewski, and L. E. Nacke, "The gamification user types Hexad scale," *CHI PLAY 2016 - Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, no. October, pp. 229–243, 2016.

[36] G. F. Tondello and L. E. Nacke, "Towards customizing gameful systems by gameful design elements," *CEUR Workshop Proceedings*, vol. 2089, pp. 102–110, 2018.

[37] W. J. Boendermaker, S. S. Maceiras, M. Boffo, and R. W. Wiers, "Attentional bias modification with serious game elements: Evaluating the shots game," *JMIR Serious Games*, vol. 4, no. 2, pp. 1–10, 2016.

[38] J. F. Vermeir, M. J. White, D. Johnson, G. Crombez, and D. M. van Ryckeghem, "The effects of gamification on computerized cognitive training: Systematic review and meta-analysis," *JMIR Serious Games*, vol. 8, no. 3, 2020.

[39] S. C. Boyle, A. M. Earle, J. W. LaBrie, and D. J. Smith, "PNF 2.0? Initial evidence that gamification can increase the efficacy of brief, web-based personalized normative feedback alcohol interventions," *Addictive Behaviors*, vol. 67, pp. 8–17, 2017.

[40] R. Stevens, K. Polk, C. Merrill, F. Feng, M. Weiss, E. Brosnan, M. S. Gerber, and L. E. Barnes, "User experience design to enhance the effectiveness of mobile technologies for the treatment of mental health," *2018 Systems and Information Engineering Design Symposium, SIEDS 2018*, pp. 135–140, 2018.

[41] P. Lindner, A. Miloff, W. Hamilton, L. Reuterskiöld, G. Andersson, M. B. Powers, and P. Carlbring, "Creating state of the art, next-generation Virtual Reality exposure therapies for anxiety disorders using consumer hardware platforms: design considerations and future directions," *Cognitive Behaviour Therapy*, vol. 46, no. 5, pp. 404–420, 2017.

[42] S. W. Ha and J. Kim, "Designing a Scalable, Accessible, and Effective Mobile App Based Solution for Common Mental Health Problems," *International Journal of Human-Computer Interaction*, vol. 36, no. 14, pp. 1354–1367, 2020.

[43] A. A. Dargél, E. Mosconi, M. Masson, M. Plaze, F. Taieb, C. von Platen, T. P. Buivan, G. Pouleriguen, M. Sanchez, S. Fournier, P. M. Lledo, and C. Henry, "Toi Même, a mobile health platform for measuring bipolar illness activity: Protocol for a feasibility study," *JMIR Research Protocols*, vol. 9, no. 8, 2020.

[44] S. Michie, M. M. van Stralen, and R. West, "The behaviour change wheel: A new method for characterising and designing behaviour change interventions," *Implementation Science*, vol. 6, no. 1, p. 42, 2011.

[45] B. J. Fogg, "A behavior model for persuasive design," *Proceedings of the 4th international Conference on Persuasive Technology*, pp. 1–7, 2009.

[46] E. L. Deci and R. M. Ryan, "Self-determination theory and the facilitation of intrinsic motivation," *American Psychologist*, vol. 55, no. 1, pp. 68–78, 2000.

[47] S. Litvin, R. Saunders, M. A. Maier, and S. Lüttke, "Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: A randomized controlled trial," *PLoS ONE*, vol. 15, no. 9 September 2020, 2020.

[48] E. A. Locke and G. P. Latham, "Building a practically useful theory of goal setting and task motivation: A 35-year odyssey," *American Psychologist*, vol. 57, no. 9, pp. 705–717, 2002.

[49] S. Michie, M. Richardson, M. Johnston, C. Abraham, J. Francis, W. Hardeman, M. P. Eccles, J. Cane, and C. E. Wood, "The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions," *Annals of Behavioral Medicine*, vol. 46, no. 1, pp. 81–95, 2013.

[50] L. V. Eckerstorfer, N. K. Tanzer, C. Vogrincic-Haselbacher, G. Kedia, H. Brohmer, I. Dinslaken, and K. Corcoran, "Key elements of mhealth interventions to successfully increase physical activity: Meta-regression," *JMIR mHealth and uHealth*, vol. 6, no. 11, pp. 1–18, 2018.

[51] L. K. B. Eldredge, C. M. Markham, R. A. Ruiter, M. E. Fernández, and G. S. Kok, G.,Parcel, *Planning health promotion programs: an intervention mapping approach*. John Wiley Sons, 2016.

[52] C. Shearer, "The CRISP-DM model: the new blueprint for data mining," *Journal of data warehousing*, vol. 5, no. 4, pp. 13–22, 2000.

[53] M. Hearst, "What Is Text Mining?," *SIMS, UC Berkeley*, vol. 5, 2003.

[54] Python, "Python." https://www.python.org/, 2021.

[55] G. Diaz, "Dutch stopwords collection." https://raw.githubusercontent.com/stopwords-iso/stopwords-nl/master/stopwords-nl.txt, 2016.

[56] GameBus, "GameBus." https://www.gamebus.eu/, 2021.

[57] R. Nuijten, P. Van Gorp, A. Khanshan, P. Le Blanc, P. Van den Berg, A. Kemperman, and M. Simons, "Evaluating the impact of personalized goal setting on engagement levels of government staff with a gamified mHealth tool: results from a two-month randomized intervention trial (Preprint)," *JMIR mHealth and uHealth*, 2021.

[58] S. R. Evans, "Clinical trial structures," *Journal of Experimental Stroke and Translational Medicine*, vol. 3, no. 1, pp. 8–18, 2010.

[59] J. van Heugten and L. James, "Investigating the impact of personalized treatment goals on engagement levels of SMI patients with an mHealth tool: a pilot study." https://doi.org/10.6084/m9.figshare.20226135, 2022.

[60] J. L. Reynolds, "Intrinsic Motivation Inventory (IMI)," *Handbook of Research on Electronic Surveys and Measurements*, no. Imi, pp. 1–12, 1984.

[61] M. B. Donnellan, F. L. Oswald, B. M. Baird, and R. E. Lucas, "The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality," *Psychological Assessment*, vol. 18, no. 2, pp. 192–203, 2006.

[62] R Core Team, "R: A language and environment for statistical computing.." http://www.r-project.org/index.html, 2020.

[63] M. Csikszentmihalyi, "Flow: The psychology of optimal experience," *Harper Row New York*, vol. 1990, 1990.

[64] P. Lindgreen, K. Lomborg, and L. Clausen, "Patient use of a self-

monitoring app during eating disorder treatment: Naturalistic longitudinal cohort study," *Brain and Behavior*, vol. 11, no. 4, pp. 1–12, 2021.

[65] G. Pramana, B. Parmanto, J. Lomas, O. Lindhiem, P. C. Kendall, and J. Silk, "Using mobile health gamification to facilitate cognitive behavioral therapy skills practice in child anxiety treatment: Open clinical trial," *JMIR Serious Games*, vol. 20, no. 5, 2018.

[66] T. A. Dennis and L. J. O'Toole, "Mental health on the go: Effects of a gamified attention-bias modification mobile application in trait-anxious adults," *Clinical Psychological Science*, vol. 2, no. 5, pp. 576–590, 2014.

Appendix A
Overview of personalized tasks

TABLE III: LIST OF PERSONALIZED TASKS FOR EACH PARTICIPANT.

Participant	Description of personalized tasks	Importance factor	Frequency	Max. number of points per week	Points
1	Poets vandaag 2 keer je tanden	3	2 × per day	+17.5	+1
	Maak zondag de woonkamer en keuken schoon	3	1 × per week	+17.5	+18
	Nuttig deze week 5 keer een gezonde maaltijd	3	5 × per week	+17.5	+4
	Maak vrijdag het toilet schoon en stofzuig de kamer	3	1 × per week	+17.5	+18
	Eet vanavond 4 boterhammen in plaats van 6 boterhammen	3	1 × per day	+17.5	+3
	Maak zelfstandig een wandeling van 25 minuten	3	1 × per week	+17.5	+18
2	Doe vandaag de was: wassen, ophangen en opvouwen	3	1 × per 2 days	+24.2	+7
	Haal op woensdagochtend en donderdagochtend de boodschappen	3	2 × per week	+24.2	+12
	Ga deze week 2 keer in de ochtend naar de dagbesteding	1	2 × per week	+8.1	+4
	Ga deze week 2 keer een wandeling maken van een uur	3	2 × per week	+24.2	+12
	Ga samen met [X] een half uur fietsen [1]	3	1 × per week	+24.2	+24
3	Maak een wandeling van 30 minuten	1	1 × per day	+17.5	+3
	Ga om 22:30 uur naar bed en doe om 23:00 uur het licht uit	3	1 × per day	+52.5	+8
	Doe in de ochtend een leuke activiteit van 60 minuten	2	1 × per day	+35.0	+5
4	Blaas op de kornet voor 15 minuten	3	1 × per day	+21.0	+3
	Ga samen met [X] een wandeling maken [1]	3	1 × per day	+21.0	+3
	Eet een stuk fruit	3	1 × per day	+21.0	+3
	Speel vandaag een spelletje	3	1 × per day	+21.0	+3
	Lees deze middag in een boek voor 30 minuten	3	1 × per day	+21.0	+3

[1] [X]: Name has been removed for privacy reasons.

Appendix B
Overview of non-personalized tasks

TABLE IV: LIST OF NON-PERSONALIZED TASKS FOR EACH PARTICIPANT.

Participant	Description of non-personalized tasks	Importance factor	Frequency	Max. number of points per week	Points
All	Ga voor een korte wandeling, fietstocht of boodschap	3	1 × per day	+17.5	+3
	Eet op tijd een gezonde maaltijd	3	1 × per day	+17.5	+3
	Eet een stuk fruit vandaag	3	1 × per day	+17.5	+3
	Poets vandaag je tanden	3	1 × per day	+17.5	+3
	Geef jezelf een compliment	3	1 × per day	+17.5	+3

Appendix C
Pre-Test Survey

TABLE V: DETAILS OF PRE-TEST SURVEY.

#	Question or statement	Type
Sample Demographics		
1.	What is your e-mail address?	Open
2.	What is your age?	One of: 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, or 65 or older
3.	What is your gender?	One of: Female, Male, or Other
Intrinsic Motivation Inventory: *Interest/Enjoyment* [1]		

TABLE V: DETAILS OF PRE-TEST SURVEY (CONTINUED).

#	Question or statement	Type
4.	I think I will really enjoy using the GameBus web application.	5-point Likert
5.	The GameBus web application seems fun to use.	5-point Likert
6.	I think the GameBus web application will be boring. (R)	5-point Likert
7.	The GameBus web application won't hold my attention at all. (R)	5-point Likert
8.	I would describe the GameBus web application as very interesting.	5-point Likert
9.	I think the GameBus web application could be quite enjoyable.	5-point Likert
Intrinsic Motivation Inventory: *Perceived Competence* [2]		
10.	I think I will be pretty good at using the GameBus web application.	5-point Likert
11.	I think I will be pretty good at using the GameBus web application, compared to other participants.	5-point Likert
12.	I think I will be satisfied with my performance in the GameBus web application.	5-point Likert
13.	I will be pretty skilled at using a web application.	5-point Likert
14.	Using the GameBus web application is something I will not do very well, I guess. (R)	5-point Likert
Intrinsic Motivation Inventory: *Perceived Choice* [3]		
15.	I believe I have some choice in using the GameBus web application.	5-point Likert
16.	I feel like it is not my own choice to use the GameBus web application. (R)	5-point Likert
17.	I don't really have a choice to use the GameBus web application. (R)	5-point Likert
18.	I feel like I have to participate. (R)	5-point Likert
19.	I'm going to use the GameBus web application because I don't have a choice. (R)	5-point Likert
20.	I'm going to use the GameBus web application because I want to.	5-point Likert
21.	I'm going to use the GameBus web application because I have to. (R)	5-point Likert
Intrinsic Motivation Inventory: *Pressure/Tension* [4]		
22.	I don't feel nervous. (R)	5-point Likert
23.	I feel very tense, because I am going to work with the GameBus web application.	5-point Likert
24.	I'm very relaxed about using the GameBus web application. (R)	5-point Likert
25.	I am anxious about using the GameBus web application.	5-point Likert
26.	I feel pressure while using the GameBus web application.	5-point Likert
mini-IPIP: *Extraversion* [6]		
27.	In general, I am the life of the party.	5-point Likert
28.	In general, I do not talk a lot.	5-point Likert
29.	In general, I talk to a lot of different people at parties.	5-point Likert
30.	In general, I keep in the background.	5-point Likert
mini-IPIP: *Agreeableness* [6]		
31.	In general, I sympathize with others' feelings.	5-point Likert
32.	In general, I am not interested in other people's problems.	5-point Likert
33.	In general, I feel others' emotions.	5-point Likert
34.	In general, I am not really interested in others.	5-point Likert
mini-IPIP: *Conscientiousness* [5]		
35.	In general, I get chores done right away	5-point Likert
36.	In general, I often forget to put things back in their proper place.	5-point Likert
37.	In general, I like order.	5-point Likert
38.	In general, I make a mess of things.	5-point Likert
mini-IPIP: *Neuroticism* [6]		
39.	In general, I have frequent mood swings.	5-point Likert
40.	In general, I am relaxed most of the time.	5-point Likert
41.	In general, I get upset easily	5-point Likert
42.	In general, I seldom feel blue.	5-point Likert
mini-IPIP: *Openness to experience* [6]		
43.	In general, I have a vivid imagination.	5-point Likert
44.	In general, I Am not interested in abstract ideas.	5-point Likert

TABLE V: DETAILS OF PRE-TEST SURVEY (CONTINUED).

#	Question or statement	Type
45.	In general, I have difficulty understanding abstract ideas.	5-point Likert
46.	In general, I do not have a good imagination.	5-point Likert

[1] Questions 4 to 9 were displayed in random order. [2] Questions 10 to 14 were displayed in random order. [3] Questions 15 to 21 were displayed in random order. [4] Questions 22 to 26 were displayed in random order. [5] Questions 27 to 46 were displayed in random order.

APPENDIX D
INTERMEDIATE-TEST AND POST-TEST SURVEYS

TABLE VI: DETAILS OF INTERMEDIATE-TEST AND POST-TEST SURVEYS.

#	Question or statement	Type
Sample Demographics		
1.	What is your e-mail address?	Open
Intrinsic Motivation Inventory: *Interest/Enjoyment* [1]		
2.	I enjoyed using the GameBus web application very much.	5-point Likert
3.	The GameBus web application was fun to use.	5-point Likert
4.	I thought using the GameBus web application was boring. (R)	5-point Likert
5.	The GameBus web application did not hold my attention at all. (R)	5-point Likert
6.	I would describe the GameBus web application as very interesting.	5-point Likert
7.	I thought the GameBus web application was quite enjoyable.	5-point Likert
8.	While I was using the GameBus web application, I was thinking about how much I enjoyed it.	5-point Likert
Intrinsic Motivation Inventory: *Perceived Competence* [2]		
9.	I think I am pretty good at using the GameBus web application.	5-point Likert
10.	I think I am quite good at using the GameBus web application, compared to other participants.	5-point Likert
11.	I am satisfied with my performance in the GameBus web application.	5-point Likert
12.	I was pretty skilled in using the GameBus web application.	5-point Likert
13.	Using the GameBus web application is something I couldn't do very well. (R)	5-point Likert
14.	After using the GameBus web application for awhile, I felt pretty competent.	5-point Likert
Intrinsic Motivation Inventory: *Perceived Choice* [3]		
15.	I believe I had some choice in using the GameBus web application.	5-point Likert
16.	I felt like it was not my own choice to use the GameBus web application. (R)	5-point Likert
17.	I didn't really have a choice to use the GameBus web application. (R)	5-point Likert
18.	I felt like I had to participate. (R)	5-point Likert
19.	I used the GameBus web application because I had no choice. (R)	5-point Likert
20.	I used the GameBus web application because I wanted to.	5-point Likert
21.	I used the GameBus web application because I had to. (R)	5-point Likert
Intrinsic Motivation Inventory: *Pressure/Tension* [4]		
22.	I did not feel nervous at all while I was using the GameBus web application. (R)	5-point Likert
23.	I felt very tense while I using the GameBus web application.	5-point Likert
24.	I was very relaxed in using the GameBus web application. (R)	5-point Likert
25.	I am anxious while using the GameBus web application.	5-point Likert
26.	I felt pressure while using the GameBus web application.	5-point Likert

[1] Questions 2 to 8 were displayed in random order. [2] Questions 9 to 14 were displayed in random order. [3] Questions 15 to 21 were displayed in random order. [4] Questions 22 to 26 were displayed in random order.

APPENDIX E
POST INTERVIEW QUESTIONS

TABLE VII: DETAILS OF POST-INTERVIEW.

#	Question
GameBus web application	
1.	What did you think of the GameBus web application? - What did you like and dislike about the web application? - Did you enjoy using the web application?
2.	What did you think of the functionalities within the web application? (e.g. points, leaderboard, Joe)
3.	What kind of functionalities would you like to see more in the web application?
4.	What still needs to change in the web application in order to implement it within GGZ Centraal?
5.	How easy was it to use the GameBus web application? (e.g. logging in, completing challenges)
6.	What grade (1-10) would you give the GameBus web application?
Personalized tasks (A) vs. Non-personalized tasks (B)	
7.	What did you think of tasks A? - What did you like and dislike about these tasks? - Did you enjoy these tasks?
8.	What did you think of tasks B? - What did you like and dislike about these tasks? - Did you enjoy these tasks?
9.	Which tasks do you prefer (A or B)? Why?
Motivation	
10.	Did the web application motivate you to accomplish tasks? If so, how?
11.	Did the web application motivate you to be more actively complied with your treatment program? If so, how?
12.	Do you think GameBus is a motivational web application to use in your treatment program?
13.	What is your main reason for using the GameBus web application?
14.	What helps you to motivate?
Other	
15.	Any further questions or comments?

APPENDIX F
BIGRAMS AND TRIGRAMS

TABLE VIII: BIGRAMS AND TRIGRAMS OF TREATMENT PLANS.

Index	Bigram	Occurence	Index	Trigram	Occurence
1	ik wil	4720	1	ik wil graag	685
2	ik heb	2531	2	ik heb een	640
7	ik ga	1076	3	ik wil mijn	437
9	ik ben	893	6	ik wil een	311
14	ik kan	674	11	ik neem mijn	203
24	ik bespreek	515	15	ik wil leren	193
31	ik gebruik	436	17	ik wil meer	185
33	ik krijg	411	20	ik wil minder	158
34	ik neem	405	21	ik bespreek met	152
36	wil ik	396	25	ik ga naar	142
42	acties ik	369	40	ik wil weer	112
43	actie ik	361	41	ik gebruik mijn	111
44	heb ik	360	45	ik maak gebruik	103
			46	ik bespreek mijn	103
			47	ik wil mij	102
			48	ik heb mijn	101