Computers in Human Behavior Reports 17 (2025) 100586

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Full length article

PENguIN: A mental health application employing gamification and token economy to boost therapeutic adherence in young users

Marco Cremaschi [a,*], Giulia Rosemary Avis [a], An Qi Zhao [a], Elia Guarnieri [a], Anna Panzeri [b], Andrea Spoto [b]

[a] Department of Informatics, Systems and Communication, University of Milano-Bicocca, viale Sarca 336, Milan, 20126, Italy
[b] Department of General Psycology, University of Padua, via Venezia 8, Padua, 35131, Italy

ARTICLE INFO

Keywords:
Cognitive behavioral therapy
Digital interventions
Gamification
mHealth
Mental health
Token economy
Treatment adherence
Usability

ABSTRACT

This paper explores the digitalisation of token economy principles within the context of mental health care, focusing on the PEnguIN application. As part of the Digital Intervention in Psychiatric and Psychological Services (DIPPS) project, the app aims to enhance therapeutic adherence among young patients experiencing Ultra-High Risk and First Episode Psychosis. It employs gamification elements, avatars, and a token economy system to foster engagement and self-expression. The study presents preliminary usability testing, employing a mixed-methods approach to assess the app's acceptability and identify potential issues. The findings provide valuable insights for future developments in the digitisation of psychodiagnostic and psychiatric tools, contributing to innovative mental health therapy approaches through technology.

1. Introduction

The complex relationship between mental health and human well-being is a focus of extensive scientific research. Within this realm, achieving psychological stability is a massive endeavour, fraught with complexities and uncertainties. In 2019, the global landscape witnessed a staggering prevalence of mental disorders, with the WHO estimates revealing an alarming statistic: nearly one billion individuals struggled with the numerous challenges posed by conditions such as anxiety, depression, and schizophrenia (World Health Organization, 2016a). The latter, in particular, has emerged as an important area of research due to its major impact on cognitive functions and behavioural patterns. The WHO's assessment, demonstrating its incidence at about one in every 300 adults, emphasised the need to address this severe psychiatric condition.[1]

In recent years, there has been a notable surge in the development and adoption of innovative digital technologies within healthcare organisations globally (World Health Organization, 2016b). The term "Digital Health" refers to a wide range of technologies used for patient care, as well as the collecting and transmission of health data. Prominent components of Digital Health include mobile applications, wearable devices, big data analytics, and Artificial Intelligence (AI) (Borrelli & Ritterband, 2015; Dattani, Ritchie, & Roser, 2021).

While Digital Health has enormous potential, widespread adoption confronts various obstacles, and the full range of its capabilities has yet to be completely achieved. On a smaller scale, these technologies have been demonstrated to have beneficial effects and to be feasible and efficient for prompt digital treatments (Hjorthøj, Stürup, McGrath, & Nordentoft, 2017). In psychiatry, digital health solutions improve the overall quality of life by detecting and controlling dysfunctional behaviours and symptoms (Firth & Torous, 2015), urging people to adopt healthier habits, which contributes to a longer life expectancy for people with psychiatric disorders (Hjorthøj et al., 2017).

In response to this need, we have developed the Digital Intervention in Psychiatric and Psychologist Services (DIPPS) project, a multi-channel platform for 24/7 patient monitoring, aiming to reduce the risk of exacerbating psychiatric conditions. This system consists of two interconnected components:

1. PsiTools (Psychological Tools): a centralised hub for managing the entire therapeutic journey, facilitating patient progress and treatment plans oversight comprised of different features, including a predictive model and decision support system, which is engineered to aid mental health professionals in making informed diagnoses and decisions.
2. PEnguIN (Psychosis Early INtervention): a mobile application empowering patients with access to therapy-related information,

* Corresponding author.
E-mail address: marco.cremaschi@unimib.it (M. Cremaschi).

[1] World Health Organization (WHO), mental disorder report: www.who.int/news-room/fact-sheets/detail/mental-disorders.

https://doi.org/10.1016/j.chbr.2025.100586
Received 1 August 2024; Received in revised form 10 December 2024; Accepted 1 January 2025
Available online 6 January 2025

a chat-bot function which enables instant communication with therapists, gamification features and a token economy system to actively engage the patient in therapy tasks and assignments.

This paper focuses on the PEnguIN app, designed for Ultra-High Risk (UHR) and First Episode Psychosis (FEP) adolescents and young adults. The application simplifies assessment administration and promotes therapeutic goals achievement with Cognitive Behavioural Therapy (CBT) tools (Wenzel, 2017). It supports users in completing CBT assignments and performing strategies (e.g., Dialectical Behaviour Therapy skills) at high-stress moments, as identified by physiological monitoring devices. Furthermore, it incorporates an avatar to represent users, providing a safe space for user expression, enhancing immersion, and stimulating creativity within the app.

This study is part of a larger process of building a digitised Token Economy (TE) and gamification system to maximise therapeutic adherence and reduce dropout risk. In a previous study, a framework, originally developed for implementing a TE system in an inpatient setting, was adapted to establish a predefined procedure for therapists to create and customise the system (the citation is momentarily unavailable due to anonymisation guidelines). Following this first version, after feedback and commentary, some modifications were made to test the usability of the solution on users. Specifically, the PEnguIN application prototype was tested with 20 participants within the target user age range. Additional details on the participants are given in the Table A.3. Before proceeding with the intended clinical trial, the evaluation was conducted to pinpoint any issues related to the designed therapeutic features, TE and gamification elements, as well as aspects concerning navigation, aesthetics, content clarity, error handling, and user satisfaction. The aim was to gather qualitative and quantitative user feedback to guide iterative adjustments to the PEnguIN application, enhancing its usability and overall acceptability. This study is not merely a validation of existing tools but an effort in digitising already used physical therapeutic tools for mental health. The findings and lessons learned from this research offer valuable insights and guidelines for future endeavours in the digitisation of psychodiagnostic and psychiatric tools, as well as the enhanced adherence to treatments such as CBT with the use of digitised Token Economy and gamification systems.

The article is organised as follows: in Section 2, basic concepts are presented to understand the objectives and functionalities of the application studied in this paper. In Section 3, related works are reviewed, providing context and highlighting relevant research in the field. Section 4 offers a detailed description of the PEnguIN application, outlining its functionalities. In Section 5, the experimental work is outlined. Section 6 presents the results of the study. A discussion on the obtained results and future work is provided in Section 7.

2. Preliminaries

The Section will delve into the context surrounding the design and development of the application, with particular attention to the target users and the research informing both the TE system and gamification features.

2.1. Ultra-high risk and first episode psychosis

Psychosis differs from other mental health illnesses in that it develops gradually and subtly. Indeed, it eludes identification because it remains hidden and unrecognised for long periods, sometimes years. The disorder's concealed nature exacerbates the distress experienced by affected people, allowing it to prolong and deteriorate without appropriate care (Addington, Farris, Devoe, & Metzak, 2020; Walker et al., 2013).

Experiencing psychosis personally includes a series of significant and devastating changes (Fusar-Poli et al., 2022). Individuals may experience heightened fear, confusion, and a deep disconnection from reality. Physically and perceptually, their perspective of the environment may be distorted and unpredictable, making even basic everyday chores difficult. Socially, psychosis often leads people to withdraw from their social networks, resulting in feelings of isolation and loneliness. Regarding cognition, it frequently interferes with thought processes, decision-making ability, and maintaining a coherent self-identity.

Psychotic disorders often appear early in childhood and proceed to chronic problems, with schizophrenia being a classic example. This highlights the vital need for early intervention, as such interventions cannot only reduce the chance of developing psychosis but also significantly impact the illness's progression. Schizophrenia stands as the most severe form within the spectrum of psychotic disorders (Arciniegas, 2015). Patients have a much lower quality of life and life expectancy than the normal population, estimated to be 15–20 years shorter (Hjorthøj et al., 2017).

Several factors contribute to the reduced life expectancy associated with schizophrenia. One primary contributor is the correlation between the illness and various other health issues, including obesity, the early onset of cardiovascular disease, self-destructive and suicidal behaviours, and recurrent substance abuse leading to addiction (Hjorthøj et al., 2017). These comorbidities, when paired with the basic symptoms of schizophrenia, increase the total health risks experienced by those affected. Recognising these connected elements is critical for establishing comprehensive treatment and support programs to promote a healthy lifestyle. Additionally, individuals with schizophrenia face considerable interpersonal challenges, such as a diminished capacity to form close relationships, lower success rates in obtaining and maintaining employment (Bouwmans, de Sonneville, Mulder, & Hakkaart-van Roijen, 2015), reduced participation in social activities, and generally inadequate social functioning (Bellack, Bennett, Gearon, Brown, & Yang, 2006).

This condition significantly strains the healthcare system, requiring extended medical support involving pharmacological, psychological, and social interventions. Consequently, it considerably impacts healthcare expenditures and resource allocation across different countries. For instance, a recent economic study undertaken in the United Kingdom revealed the potential for significant financial savings, totalling £125 million over three years, with the implementation of early and preventative treatment strategies (Bucci et al., 2015).

Before the onset of the full-blown disease, adolescents and young adults may progress through distinct phases marked by prodromic or sub-threshold symptoms, referred to as UHR and FEP (Arciniegas, 2015). These phases function as early indicators of more severe psychotic disorders, including schizophrenia, as discussed previously, and hold significant relevance in both mental health research and clinical practice.

Developing effective approaches and procedures for recognising situations of risk related to UHR and FEP is vital for encouraging and facilitating early intervention (Ologundudu et al., 2023). These efforts aim to alleviate the pain and misery experienced by individuals in these vulnerable populations, as well as reduce the significant economic impact that persistent psychotic illnesses have on healthcare systems and society as a whole. Such initiatives can improve mental healthcare by increasing efficiency and cost-effectiveness while improving the well-being of those afflicted. Individuals can re-establish a sense of stability and normalcy by receiving support, education, and therapeutic interventions, thereby minimising the life-altering impacts of psychosis (Meneghelli et al., 2013).

We aim to enhance patients' understanding of their disease and assist therapists in the complex process of diagnosis and evidence collection throughout the therapy process. To achieve these objectives, a gamified experience and a TE system were developed to motivate patients to engage with the application consistently, increasing the effectiveness of therapy.

2.2. Gamification

In the context of therapeutic interventions for UHR and FEP patients, who often face difficulties in adhering to treatment regimens (Killikelly, He, Reeder, Wykes, et al., 2017), the incorporation of gamification and TE mechanisms within a mobile application emerges as a promising avenue (Firth & Torous, 2015). This pairing can potentially increase treatment adherence by combining the complexities of mental health care with the appealing attraction of gaming theories.

Gamification is the deliberate incorporation of game design components and principles into non-game environments, with the goal of increasing engagement, motivation, and behavioural regulation in sectors such as education, healthcare, business, and personal development (Deterding, Dixon, Khaled, & Nacke, 2011). It draws on the attraction of games to promote desirable actions and outcomes, transforming ordinary tasks into engaging experiences enriched with play mechanics with features like points, badges, levels, and challenges. These design elements appeal to basic human drives, persuading users to devote time and effort to the experience. The interplay between play and therapy has garnered considerable attention, shedding light on play's psychological and philosophical dimensions and its therapeutic implications. The notion of "flow" encapsulates a state of complete immersion in an activity, frequently linked with play-related efforts, in which individuals are fully engrossed in the task at hand, oblivious to unrelated distractions (Csikszentmihalyi, 1975; Csikszentmihalyi & Csikzentmihaly, 1990). A "gameful" approach to life echoes the concept of flow, emphasising the transformative potential of play in addressing real-world challenges and enhancing well-being (McGonigal, 2011, 2016). A work in the state of the art underscores the therapeutic value of play, providing individuals with a space for creative expression and psychological transformation (Winnicott, 2005). Transitional phenomena, such as play and creativity, serve as bridges between internal fantasies and external realities, facilitating self-discovery and healing.

Gamification emerges as a viable strategy in eHealth to boost well-being and deal with mental health difficulties. Several studies have investigated the usefulness of gamified therapy, notably in engaging young people and alleviating psychiatric diseases such as Attention Deficit Hyperactivity Disorder (ADHD), autism spectrum disorders, anxiety, and dyslexia (Stephenson et al., 2023; Zayeni, Raynaud, & Revet, 2020). Similarly, the incorporation of TE principles offers potential as a therapeutic technique for improving treatment adherence and generating positive behavioural change in people dealing with mental illnesses.

2.3. Token economy in cognitive behavioural therapy

Within mental health treatments, psychoeducation and CBT stand out as crucial resources for clinicians, as they enable them to intervene in a way that either prevents psychosis in susceptible individuals or aids in the recovery of those who have already experienced a psychotic episode (Herrera et al., 2023). Additionally, they play a critical role in preventing the worsening of mental disorders and promoting an improved quality of life for people afflicted by them by addressing and controlling symptoms and behaviours early in the illness trajectory.

In the first case, psychoeducation represents an educational approach rooted in the field of clinical research, focused on preventing relapse and promoting health restoration among individuals with psychiatric disorders (Zhao, Sampson, Xia, & Jayaram, 2015). This can be achieved by providing them and their families with the knowledge and competencies required to collaboratively engage with mental health professionals, promote treatment and therapy compliance, and facilitate the reintegration of patients into their respective communities. Conversely, CBT aims to modify maladaptive thought patterns and behaviours to improve emotional well-being (Wenzel, 2017). Its effectiveness lies in its personalised and collaborative nature, with therapists developing interventions and homework assignments to enhance patient adherence to therapy goals (Kazantzis, Deane, Ronan, & L'Abate, 2005).

In this context, the TE intervenes as a strategy employed to increase the likelihood of treatment adherence: the patient receives tokens as a form of acknowledgement and reward for fulfilling homework requirements and advancing along the therapeutic journey. In this regard, the TE constitutes a system of positive reinforcement based on the premise that individuals are induced to take specific actions if they are associated with immediate or delayed rewards. As symbolic markers of worth, tokens bolster favourable behaviours, thus establishing an environment wherein individuals are innately driven to pursue predefined objectives (Doll, McLaughlin, & Barretto, 2013). This dynamic fosters a motivation and satisfaction cycle, promoting continuous engagement and progress. The framework offered by the TE proves to be applicable across diverse settings, spanning from educational to professional areas and beyond, including the mental health field: over the past three decades, a substantial body of literature has emerged, delving into the effectiveness and nuances of the TE approach among psychiatric inpatients programs (Glowacki, Warner, & White, 2016; Glynn, 1990).

Similarly, studies (Deo, Johnson, Kancharla, O'Horo, C and Kashyap, others, 2020) have shown the potential of instant gratification to enhance guideline adherence. In our study, instant gratification was operationalised through the digitised and gamified TE system, where patients received points for completing specific tasks. This method harnesses the principles of immediate positive reinforcement, encouraging patients to consistently follow prescribed guidelines and maintain their engagement with their treatment plans.

Considering the outcomes of the ever-evolving Digital Health and the optimistic perspective on the convergence of technology and psychiatric programs, the DIPPS project attempts to extend and digitalise the principles of the TE beyond inpatient settings to include outpatient care. Specifically, the PEnguIN mobile application is designed to stimulate the active participation of outpatients and consequently their treatment and therapy compliance: through task completion, like compiling a questionnaire or adding diary entries, outpatients can earn points tailored to their therapeutic goals, which are visually represented as a growing tree within the app (Figs. 2, 3, 4). This visual depiction offers positive reinforcement and signals progress. Furthermore, accumulated points can be used in the app gamification system to personalise avatars, enhancing engagement (the citation is momentarily unavailable due to anonymisation guidelines).

A component of the TE system digitisation process entailed establishing a predefined procedure for therapists to create and customise the system. As previously mentioned (Section 1), the proposed technological solution comprises three closely interconnected elements. Health professionals customise the TE system via the web-based PsiTools platform, while users/patients interact with the TE system through the PEnguIN application. To outline the process (Fig. 1), a framework (Ghezzi & Lewon, 2022), originally developed for implementing a TE system in person, was adapted for the study's objectives and is outlined below:

1. *Formulate clear and precise objectives and choose pertinent target responses*: as the focus is improving therapeutic adherence, clinicians can establish goals and tasks within the PEnguIN application, such as completing food diaries or psychodiagnostic questionnaires, to facilitate progress towards the therapeutic objectives.
2. *Repeatedly, accurately, and reliably assess the target behaviour*: through PsiTools, health professionals can monitor patients' progress with assigned tasks, including initiation, completion, and time taken. In face-to-face sessions, health professionals utilise these metrics to track progress over time and engage patients in discussions about their motivations.

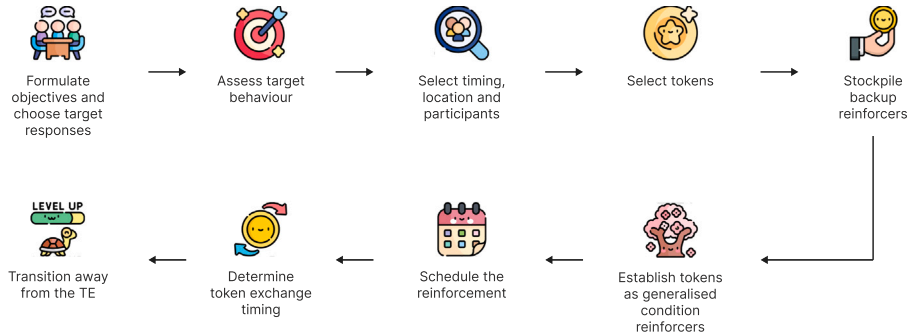


Fig. 1. Studied framework to implement the Token Economy system.

3. *Select the timing, location, and participants for the operation of the TE*: clear instructions are given to all the parts involved, and tasks are assigned digitally, with timing regulated through scheduled events. Notifications prompt task completion, but ultimately, it is up to the patient, as non-completion provides insight into progress.
4. *Select tokens*: the digital tokens are represented as coins, as several studies consider them valid alternative to physical ones (Horner, Hew, & Tan, 2018; Robacker, Rivera, & Warren, 2016; Williamson & McFadzen, 2020). The PEnguIN application enables patients and therapists to customise the tokens by selecting the imagery displayed on the face of the coin.
5. *Stockpile backup reinforcers*: the worth of a token is directly linked to the value of its associated backup reinforcement, which is a valuable item, activity, or privilege that individuals can buy with tokens. In the PEnguIN application, there are two types of backups: digital (e.g., avatar accessories) and tangible (e.g., purchasing a bicycle), with the latter obtained upon reaching specific token milestones and in agreement with the family.
6. *Establish tokens as generalised conditioned reinforcers*: the tokens are backed by various sources to enhance their effectiveness across different situations. In the PEnguIN application, the method involves linking tokens closely with backups through "stimulus pairing". Furthermore, the TE system visually represents progress through levels, symbolised by a growing cherry tree, and emphasises patient engagement through avatar interactions with a plant, metaphorically representing therapy and health.
7. *Specify the schedule of reinforcement*: token systems are used to reinforce desired behaviours by awarding tokens for achievements. Therefore, it is crucial to adjust the token circulation based on the patient's difficulties to maintain motivation and prevent discouragement. By using adaptive reinforcement schedules, such as varying the frequency or amount of tokens, therapists ensure that the rewards remain effective and motivating. This approach helps avoid frustration or learned helplessness, keeping the patient engaged and progressing throughout the therapy.
8. *Determine the timing for token exchange*: in the exchange-production schedule of the PEnguIN application, two approaches were implemented: response-based exchanges for in-app purchases and time-based exchanges for milestone rewards.
9. *Transition away from the TE*: shifting away from a TE system mirrors its initial implementation, as, ultimately, the goal is for the TE to become recreational rather than reward-focused. For this reason, several strategies are implemented to thin out the TE system, such as implementing a "level system", reducing token-specific rewards, and promoting self-monitoring (see Fig. 1).

3. Related works

The convergence of digitisation, TE systems, and gamification approaches across different fields highlights a multifaceted approach to tackling current mental health and well-being concerns. Studies in this regard emphasise the potential of technology-driven interventions to improve adherence and efficacy. For example, a study focusing on the use of mental health interventions, particularly CBT, through the incorporation of gamified design elements in mobile applications to address Depression, Anxiety, and Stress Disorders (DASD) among Arab adolescents reported a significant improvement in anxiety and depressive symptoms after using the application, indicating a positive impact on the mental health of participants (Amer, Shohieb, Eladrosy, Elbakry, & Abd Elrazek, 2023). Moreover, studies have been conducted on the digitisation in other fields, such as Williamson and McFadzen (2020), which analysed the impact assessment of TE methods on student behaviour in activities within an inclusive Canadian classroom, obtaining overall satisfaction with the TE system as a classroom management method. Another study demonstrated promise for the digitisation of TE system, namely (Robacker et al., 2016), which focused on the use of classroom management strategies, specifically Positive Behavioural Interventions and Supports (PBIS), to tackle the challenges faced by students with behavioural and emotional disorders in the academic environment.

Furthermore, there are several applications which employ gamification techniques to improve well-being and prevent mental health concerns among young and adults. For example, within a study (Stephenson et al., 2023), an online self-guided mental health training program was developed using gamification elements, which was found to be acceptable and engaging for participants. Moreover, it has also been studied in psychiatric contexts. A recent review (Zayeni et al., 2020) examined 22 applications incorporating game design elements, including serious games and commercial video games, for prevention or treatment in children and adolescents with psychiatric disorders such as ADHD, autism spectrum disorders, anxiety, and dyslexia. Leveraging gaming contexts to deliver therapeutic content or interventions has effectively engaged younger audiences drawn to digital technologies. However, despite promising clinical results, further research is needed to assess the long-term efficacy and real-world applicability of these interventions. In a recent study (Litvin et al., 2023), the effects of eQuoo[2], a mental health game that teaches psychological skills through gamification, were investigated, revealing an increase in resilience and a decrease in anxiety and depression in a student population. The study also underscored the potential of games for mental health, emphasising their effectiveness, scalability, and low-cost nature in supporting young adults (Caon et al., 2017).

[2] eQuoo - the Clinically Proven Gaming App that Improves Mental Wellbeing: psycapps.com/equoo/.

In this delicate context, catering to the specific needs of patients is crucial, given the significant variations in demographics, illness types, treatment phases, and literacy levels. Factors such as patients' health literacy and prior technology experience profoundly influence their engagement with health apps (Chan & Kaufman, 2011; Norman & Skinner, 2006). Additionally, addressing usability is critical as mentally vulnerable teenagers and young adults with cognitive difficulties may struggle to comprehend, learn, remember, and navigate app functionalities (Johansson, Gulliksen, & Lantz, 2015). Usability studies encompass a comprehensive evaluation across various dimensions, including app flexibility, operability, understandability, learnability, efficiency, satisfaction, attractiveness, consistency, and error rates (Zhang & Adipat, 2005). Recent international standards prioritise user feedback, shifting from previous approaches focused solely on assessing the software product. Additionally, recent International Standards (ISO 9241 International Organization for Standardization, 2018a and ISO 25062 International Organization for Standardization, 2006) provide invaluable guidance for conducting and reporting usability testing tailored specifically for mobile apps (Moumane, Idri, & Abran, 2016). These standards prioritise user feedback, departing from previous approaches focused solely on assessing the software product itself (Moumane et al., 2016).

Recent research in mHealth applications has concentrated on assessing their usability, introducing both novel (Zhou, Bao, Setiawan, Made Agus and Saptono, and Parmanto, others, 2019) and already validated (Zhang & Adipat, 2005) methodologies and frameworks for usability testing. For example, the mHealth App Usability Questionnaire (MAUQ) was specifically developed and validated for mHealth applications (Zhou, Bao, Setiawan, Made Agus and Saptono, and Parmanto, others, 2019). Moreover, numerous studies have emerged, focusing on the evaluation of the usability of these applications. For instance, one study utilised group sessions to assess health-technology applications (Brown III, Yen, Rojas, & Schnall, 2013). Additionally, the Healthy Jeart application, a game-based app tailored to foster knowledge, attitudes, and healthy habits among 8–16 year-olds, was designed, developed and tested through a questionnaire developed specifically for the study (Roldán-Ruiz, Merino-Godoy, Peregrín-Rubio, Yot-Dominguez, & da Costa, 2024). As usability testing methodologies evolve, so do opportunities to enhance the design and functionality of health-related applications. By focusing on user feedback and incorporating insights from diverse fields like psychology and game design, we can create more intuitive and impactful solutions. Leveraging innovative technologies further enhances our ability to develop effective solutions. Grounded in foundational research, our exploratory design and testing process aims to address the needs of young adults facing mental health challenges.

4. PEnguIN

The PEnguIN app, developed as part of the broader DIPPS project, aims to deliver digital telemedicine services in the psychological and psychiatric domain, specifically targeting teenagers and young adults struggling with mental health issues. Employing CBT tools and incorporating elements of game design, the app encourages patients to achieve therapeutic goals while minimising dropout risk (Amer et al., 2023).

The first implemented feature is the points system: for each completed task within the app, such as consulting informational content, completing diary entries, or engaging in psychoeducational activities, the system rewards users with points. Visualising the therapy journey as a progress bar indicating the achieved level and the distance to the next one, and the growth process of a blooming cherry tree, the app celebrates each milestone when users reach significant point thresholds, providing them with a sense of accomplishment and motivation to continue their engagement with the app and ultimately with the whole therapeutic process. Specifically, users start at level 1 by planting a small seed, and by earning points, it will grow into a sprout at level 2, a plant at level 3, a sapling at level 4, a tree at level 5, and finally, a blooming cherry tree at level 6 (Figs. 2, 3, 4).

In the context of the points system, users can exchange points for digital coins to use for purchasing items for their avatar in the digital realm or obtaining tangible goods or experiences in the real world. Therefore, tokens can be digital (e.g., avatar accessories) or tangible (e.g., purchasing a bicycle), with the latter obtained upon reaching specific token milestones and in agreement with the caregivers (e.g., family). These rewards serve as backup reinforcements, corresponding to items or activities that incentivise users within a TE.

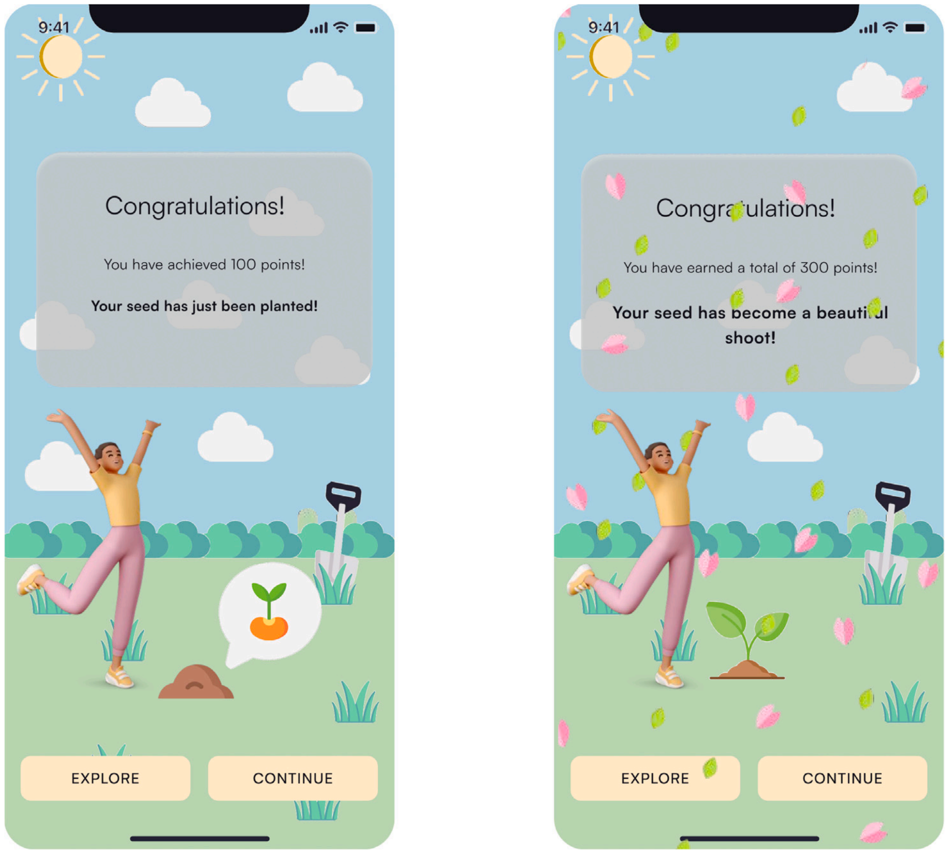
The points system feature is grounded in the crucial role of feedback, which guides users, informs their decisions and actions, and reinforces their behaviours. By providing timely and relevant feedback, users can identify their strengths and weaknesses, learn from their experiences, and adapt their behaviour accordingly. This constant feedback loop facilitates ongoing learning, skills development, and improvements, ultimately enhancing user engagement and satisfaction. Feedback comes in various forms, including positive reinforcement, constructive criticism, or corrective guidance. In this context, feedback primarily takes the form of positive reinforcement. This not only strengthens desired behaviours but also contributes to a more rewarding and engaging user experience. When users receive positive and immediate feedback for their actions or progress within the gamified environment, they experience immediate gratification and a sense of accomplishment. Furthermore, positive feedback can boost confidence, encourage active participation, and sustain high levels of user motivation in pursuing the goals of the gamified system (Aparicio, Vela, González-Sánchez, & Isla-Montes, 2012).

The aim of the points system is to leverage extrinsic motivation (Mekler, Brühlmann, Tuch, & Opwis, 2017), supported by a TE system as an advanced form of incentivisation. Participants receive tokens as recognition for their commitment and progress, convertible into benefits and rewards within and outside the application, fostering motivation and gratification.

A fundamental gamification feature of the app is its avatar, which serves as a bridge between users and health professionals while ensuring therapy confidentiality. In this scenario, the avatar is exclusively dedicated to therapy-related functions, allowing users to create representations of themselves or alternative personas for self-expression. This fosters dialogue between the user and therapist and promotes discussions on self-image (Lancini, 2020) and identity while enhancing user immersion and creativity within the app. An example can be found in Fig. 5.

The third gamification feature concerns the avatar itself: users can personalise and customise their avatars using accumulated coins, enhancing self-expression and immersion (Taylor, 2002). Research, commercial, and therapeutic settings all benefit from avatar customisation, which has been shown to increase motivation, enjoyment, involvement, and has an overall positive affect (Rahill & Sebrechts, 2021). Deliberately excluding the gender aspect from avatar design ensures gender neutrality, fostering accessibility and identity exploration freedom.

Both the points system and the avatar customisation are based on the principle of personalisation: in gamification, it enables tailoring the gamified experience to the unique preferences, skills, and specific needs of individual users. When a gamified system integrates personalised elements like custom challenges, targeted rewards, or flexible progress paths, users experience heightened engagement and motivation. Personalisation fosters a more relevant and meaningful experience by accommodating individual differences, delivering a tailored approach to optimise motivational impact. Moreover, as in the case of avatar customisation, it cultivates a sense of belonging and deeper engagement. In fact, avatar customisation, similar to real-life personalisation (e.g., mimicking friend clothes and changing hairstyles) strengthens identification with avatars. This so-called "embodied identification", which is one of the forms of avatar identification, enhances actual communication behaviour (Takano & Taka, 2022). Some studies indicate



a. Seed. *b.* Sprout.

Fig. 2. App screens illustrating the plant growth process: from seed (screen *a*) to sprout (screen *b*). The screens are presented in English to facilitate understanding. However, the application is entirely in Italian as it is developed for an Italian audience.

that in the digital world, the behaviour can affect self-esteem (Gonzales & Hancock, 2011), while the behaviour can also be shaped by various forms of digital self-representation (Yee & Bailenson, 2009). Therefore, it could be said that digital self-representation alone can influence the user's behaviour and key self-concepts, such as self-esteem (Chan, 2022).

All of these gamification features aim to engage users in activities that might not otherwise elicit the same interest or participation, transforming everyday experiences into engaging, fun, and motivating interactions.

To optimise the effectiveness of gamification features, it is essential to ensure a user-friendly experience by applying usability parameters to interface design. Beginning with the "Status" section, intimately connected to the points system, users are greeted with a captivating interface that immerses them in a virtual garden, where they can monitor the status of the symbolic cherry tree. At the centre of this interface lies an overlaying balloon containing the "Experience line" graphic, showcasing not only the accumulated score but also indicating the score required to progress to the next level. As the garden evolves gradually, the growth process becomes visible through the gradual elimination of weeds and clouds, while flowers bloom, offering a visually rewarding experience synchronised with the tree's growth (Figs. 2, 3, 4).

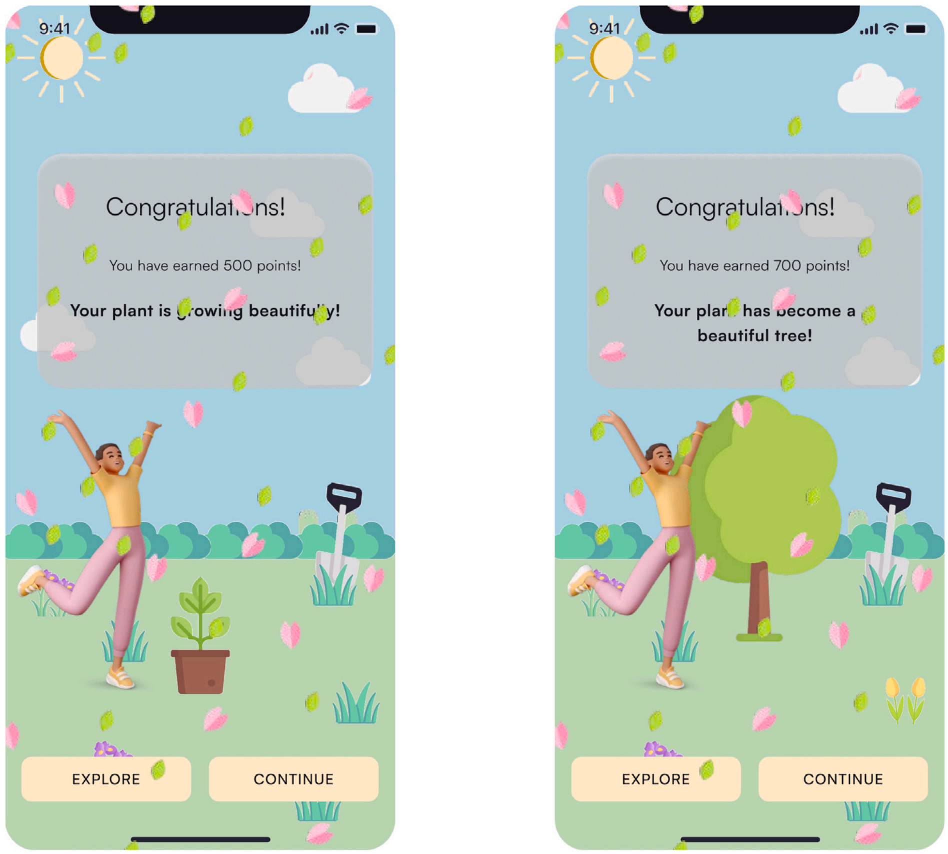
An engaging aspect of this interface is the avatar representation: as points accumulate, the avatar is depicted watering the tree, forging a direct visual connection between the user's actions and the virtual garden's growth.

Within the context of the points system, guiding users through the task completion process is crucial. Therefore, a concise informative message within the header will provide users with instructions on earning points each time they complete an exercise. Upon completing an exercise, users will be redirected to a page confirming successful completion, providing details about the earned points and those required to reach the next milestone. Feedback is essential to inform users of successful task completion, rewarding their efforts and fostering a sense of achievement. Regarding avatar customisation, the "Item" section containing various categories (tops, bottoms, shoes, hairstyles, emotions, and background) maintains a consistent structure. This interface utilises visual cues to guide users through the flow, ensuring clarity in available actions. Icons in the top bar allow users to return to the profile section or confirm avatar configuration, while the remainder of the screen is dedicated to customisation. Navigation through categories is facilitated by a horizontal scrolling navigation bar, accommodating potential expansion of customisation options. Carefully designed icons minimise the need for text, and items are presented in a spacious grid-like format, facilitating user selection.

Another significant feature within the avatar creation process is the "Store" section, offering items which are obtainable by spending digital coins earned through therapy progress. Accessible like other categories, the store section includes additional UI elements: a counter bar displays the user's available digital coins, and each item shows its price. Upon selection, a confirmation pop-up appears, and once unlocked, new items become available in their respective categories (Fig. 5).

Additionally, the app introduces various features such as:

- **Reminder and Agenda:** the calendar feature allows users to add a variety of event types, which will then appear as reminders on the homepage. These may include taking prescribed drugs, measuring physiological indicators (e.g., heartbeat), or completing established tasks (e.g., exposure activities), aiding in data



c. Plant. *d.* Little Tree.

Fig. 3. App screens illustrating the plant growth process: from plant (screen *a*) to Little tree (screen *d*). The screens are presented in English to facilitate understanding. However, the application is entirely in Italian as it is developed for an Italian audience.

collecting and improving patient adherence (Fig. 6).

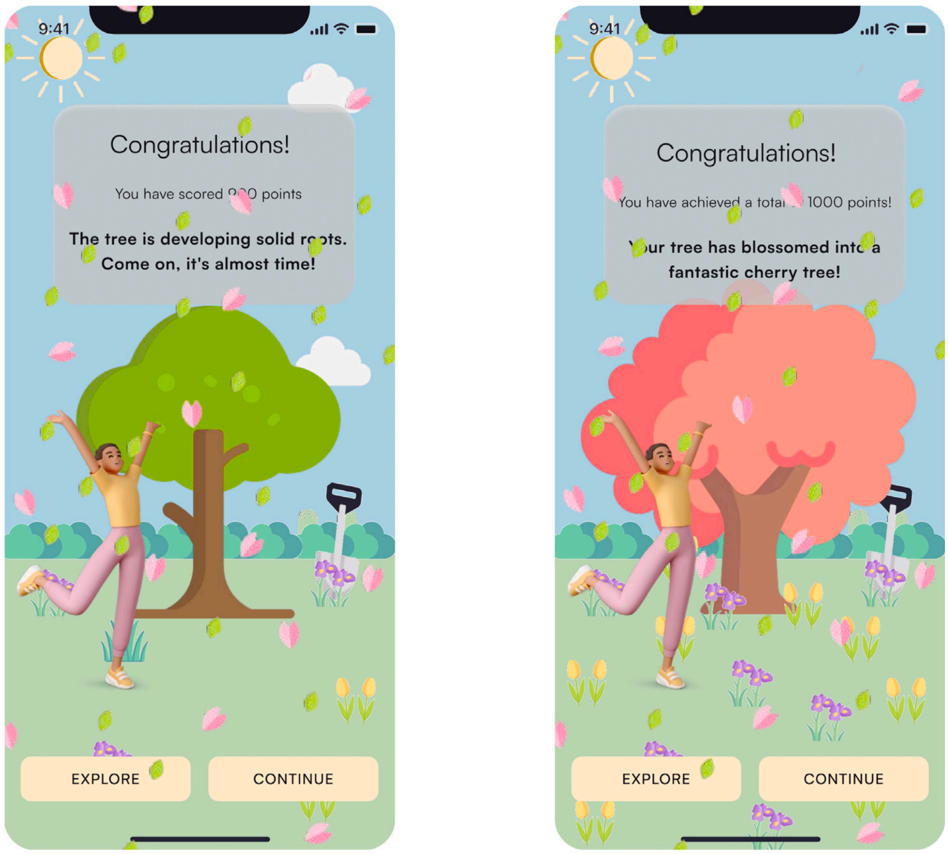
- **Psychodiagnostic tests:** patients can take psychological tests in the *assignment* section of the application. These tests mimic the structure and content of traditional paper-pencil forms, simplifying the manual data entry and scoring procedures, which are prone to high complexity and errors (Fig. 6).
- **Chatbot:** the application includes an instant chat feature that allows patients and health professionals to engage directly. In the therapist's absence, a chatbot takes over, mostly to remind the patient about scheduled events, medication intake, physiological parameter tracking, and assigned tasks (Fig. 7).

This comprehensive approach enables patients to complete assigned tasks, receive prompt feedback, and visualise their progress through gamification and a TE system developed in collaboration with their therapist. Additionally, it is important to note that the gamification and TE systems were integrated to address high dropout rates in psychiatric care. These features, previously established in traditional therapy, are now digitised to enhance engagement and interaction, especially in remote settings where maintaining motivation can be difficult. The application's design ensures that all features, including the Chatbot and reminders, are integral components of the therapeutic process, rather than extraneous variables that could interfere with the research study.

5. Evaluation

The preliminary usability testing of the PEnguIN app was designed to evaluate its suitability, identify potential design issues, and refine its features prior to clinical trials. A mixed-methods approach was adopted, combining qualitative feedback and quantitative analysis to assess usability, including task success rates, error rates, and user satisfaction. Additionally, this phase focused on non-clinical users to minimise ethical and logistical challenges associated with involving clinical populations in early-stage testing.

In planning the preliminary usability testing for the PEnguIN app, careful considerations were made regarding the selection criteria for participants. A decision was made not to directly involve individuals currently diagnosed with mental health conditions such as UHR or FEP as a preliminary usability testing was instrumental in evaluating the suitability of the app and identifying potential areas for improvement. Moreover, recruiting the clinical sample poses significant challenges obtaining informed consent and conducting clinical trials. This step will be conducted during a more advanced phase of app validation, after its features and usability have been preliminary tested. Many of these individuals may already be engaged in intensive treatment programs or require specialised care, making it ethically and logistically complex to involve them in additional preliminary research activities involving an app which has not yet reached a more polished and reliable version. Considering these reasons, recruitment efforts were redirected towards selecting a pool of 20 willing participants within



e. Tree. *f.* Cherry Blossom Tree.

Fig. 4. App screens illustrating the plant growth process: from tree (screen *e*) to blossom tree (screen *f*). The screens are presented in English to facilitate understanding. However, the application is entirely in Italian as it is developed for an Italian audience.

the age range of 18–27. This type of sample was deemed adequate and aligned with the future population of use in terms of socio-demographic characteristics. Although this sample is not clinical, the app is targeted at both clinical and non-clinical individuals undergoing prevention interventions. Moreover, individuals from the clinical population (already diagnosed as UHR or FEP) will have varying levels of functioning and impairment, including high-functioning individuals comparable to non-clinical samples like this one. Additionally, different individuals may have different areas of functional impairment, with most areas showing no impairment, thus resembling the characteristics of a non-clinical sample concerning the evaluation of usability. While the number of participants (n = 20) may appear relatively small, according to the guidelines of the Nielsen Norman Group (Nielsen, 2009), it is adequate for a preliminary study to identify issues and necessary changes in the application's usability. The inclusion criteria required participants to possess prior knowledge of CBT, either as psychology students or individuals under psychological care. This ensured that the sample comprises individuals capable of offering informed feedback on the effectiveness and utility of the CBT-related features within the app. Additionally, potential participants were required to answer a questionnaire regarding their age, smartphone proficiency and familiarity with well-being applications. The selection of this participant group was also motivated by the availability of such subjects, facilitated by university collaborations that afford easier access to both students and alumni in psychology compared to others. Prior to the usability test, participants were briefed on its objectives, procedures, and potential risks or benefits, and informed consent was provided. The usability evaluation was based on the think-aloud method (Weichbroth, 2024) to garner insights into the user-interactions with the app.

The usability testing was conducted using the Figma Proto app installed on an iPhone 13 mini, which served as the primary testing device for the participants. Additionally, a MacBook Air 13 was provided and used when the mobile prototype experienced functionality issues or participants requested a larger screen. To ensure thorough documentation, screen recording functionalities were enabled on both devices to capture interactions with the app interface and prototype.[3] Voice recording was used alongside screen recording to gather participants' feedback, observations, and commentary using the PEnguIN app.

Qualitative research methods (e.g., open-ended questions and Think Aloud methods (Charters, 2010)) were chosen alongside quantitative analysis (e.g., success task rates Nielsen & Levy, 1994) to offer a more comprehensive understanding of the usability within the PEnguIN app. The incorporation of both qualitative and quantitative research methods was driven by several considerations. Qualitative methods are invaluable for uncovering user emotions, motivations, and nuanced experiences, while quantitative methods provide measurable performance metrics that offer objective insights. The exploratory nature of the study's initial phase aimed to delve into user perceptions, behaviours, and interactions with the application. Qualitative analysis proved particularly effective in revealing insights that might be overlooked by

[3] Prototype: www.figma.com/proto/y42fkghRzcyZt81CluiXUN/.

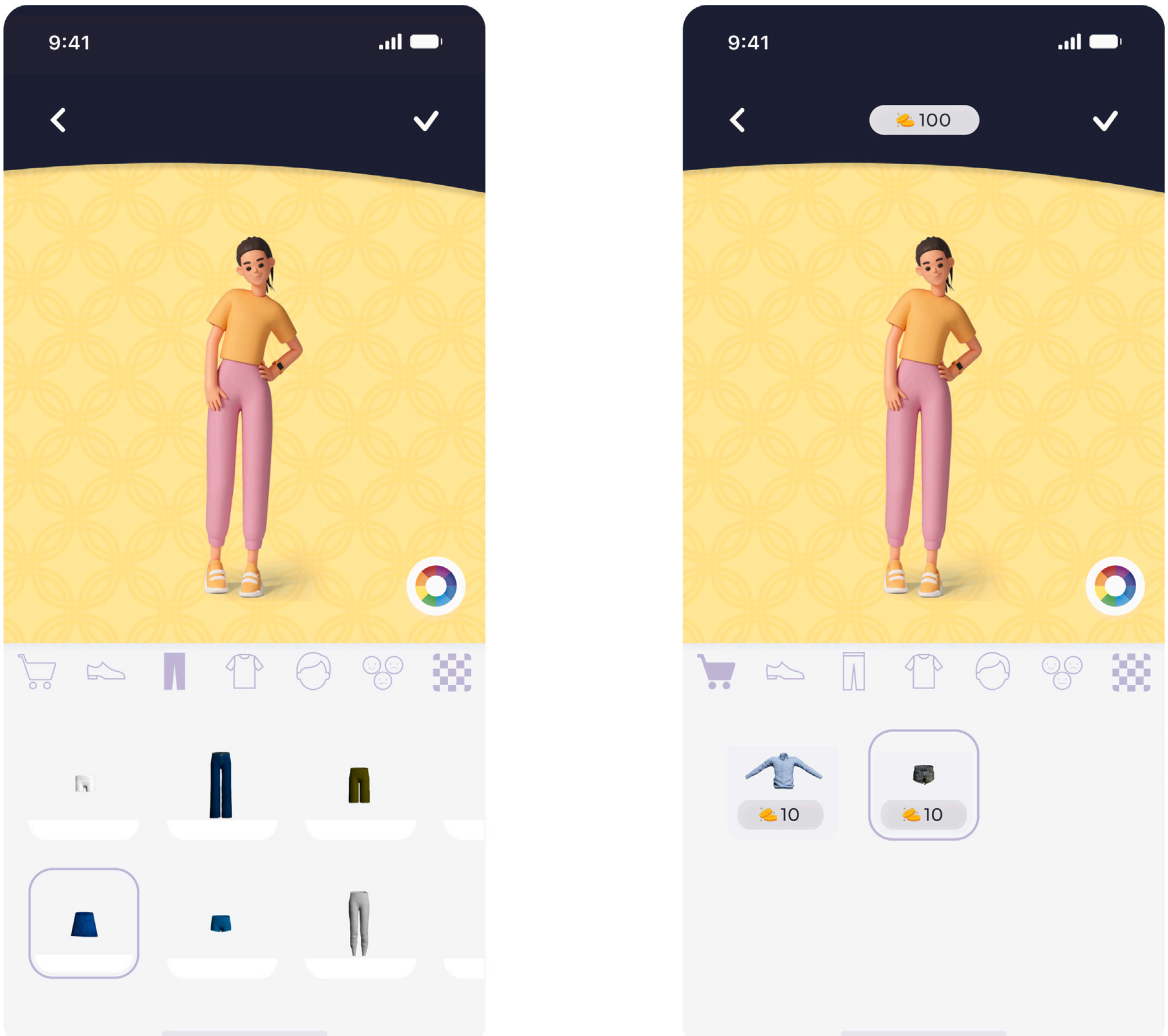


Fig. 5. Interface depicting the avatar customisation process in the Penguin app.

purely quantitative approaches. Additionally, given the user-centred design approach, understanding users' subjective experiences is vital for iterative design. Therefore, drawing from a methodological framework (Weichbroth, 2024) and a thorough review of numerous usability testing approaches, the Think Aloud technique, the System Usability Scale (SUS) questionnaire (Appendix C) and Task Succession Rate were selected for discovering usability issues within the PEnguIN application. The Think Aloud method involves the vocal description of users' cognitive processes as they engage with a certain user interface. It is widely acknowledged in usability testing and aims to capture real-time insights into user experiences (Eccles & Arsal, 2017). Additionally, during the testing phase, task succession rate was calculated, measuring effectiveness, which is defined as the accuracy and completeness with which specified users can achieve specified goals in particular environments (International Organization for Standardization, 2018b). Following the conclusion of each testing session, a feedback phase was initiated to gather detailed evaluations from participants about their experiences with the tested interface. During this, participants were asked to rate the interface's usability using the SUS questionnaire, a validated tool (Fig. C.8) for measuring user satisfaction. The SUS questionnaire consists of ten-item attitude Likert scale which assesses various aspects of the interface, including ease of use, user confidence, and satisfaction with the application. Each item is scored on a 5-point scale, and the results provide a quantitative measure of usability. The SUS score is then calculated, offering a comprehensive evaluation of the app's overall user experience (Lewis, 2018). Eventually, participants were asked to answer open-ended questions regarding their perceived satisfaction with the TE and gamification features.

During the usability testing session, we gathered several metrics to assess the application performance and identify encountered issues, including:

- *Task success rate*: participants were tasked with completing specific activities (Appendix B) within the app; a task was considered completed if participants achieved the specified goals for each task, indicating successful navigation and interaction with the application. The tasks were scored as follows: 0 for tasks not completed; 1 for tasks completed with difficulty; and 2 for tasks completed successfully. The assigned scores for the tasks were then summed and divided by the maximum possible score. This result was subsequently converted into a percentage value.

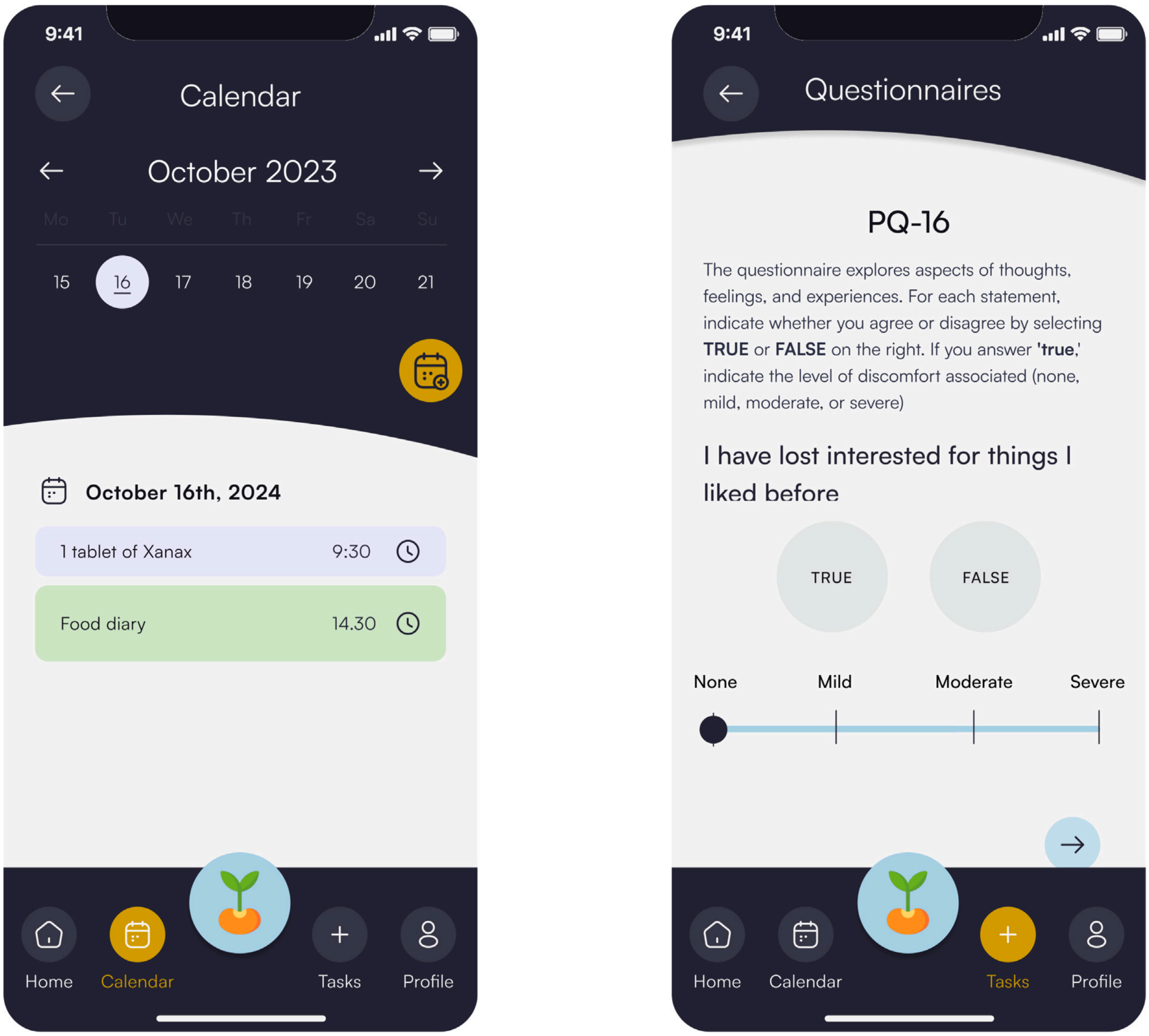


Fig. 6. Interface for the weekly calendar view and PQ16 questionnaire.

- *Errors*: participants encountered two types of errors: critical and non-critical. Critical errors occurred when participants believed they had fulfilled the task but had done so incorrectly or incompletely. Non-critical errors were recoverable instances where participants deviated from the expected pathway but successfully completed the task. An error-free rate was recorded for participants who completed tasks without encountering failures.
- *Recommendations and Feedback*: following the test session, participants were prompted to evaluate the application subjectively through the SUS questionnaire and open-ended questions.

6. Results

The study aimed to understand user perceptions of the app's usability through a mixed-methods approach, which included: effectiveness, measured by success score rate, and satisfaction, assessed using the SUS questionnaire and user feedback.

1. Demographic. The mean age of the participants was 22.7 (SD 2.94). Their gender identity was distributed as follows: 55% male, 45% female. The participant demographic profile indicated that 60% were enrolled as students, while 30% were employed and 10% were unemployed. In terms of smartphone proficiency, the majority reported a *Moderate* level (50%), followed by an *Advanced* level (30%), and *Extensive* proficiency (20%). Notably, a significant proportion of participants demonstrated *Limited* familiarity with well-being applications (45%), while 30% reported *Moderate* familiarity, and 25% reported *Extensive* familiarity. A detailed view can be found in Appendix A.

2. Effectiveness. As previously mentioned, the metrics used for evaluating the effectiveness of PEnguIN application was the success rate score, which indicates whether a particular aspect of a product needs improvement in its usability (International Organization for Standardization, 2018a).

The formula (Ferreira et al., 2023) to calculate it is shown below:

$$\text{Task success rate} = \frac{\text{number of successfully completed tasks}}{\text{total number of subtasks undertaken}} \times 100 \tag{1}$$

Participants demonstrated satisfying levels of success across all the different tasks Appendix B. The lowest success rates are on Task 4 (Accessing the progress section and score monitoring), and on Task 13 (Accessing the colour customisation mode and switching the eye

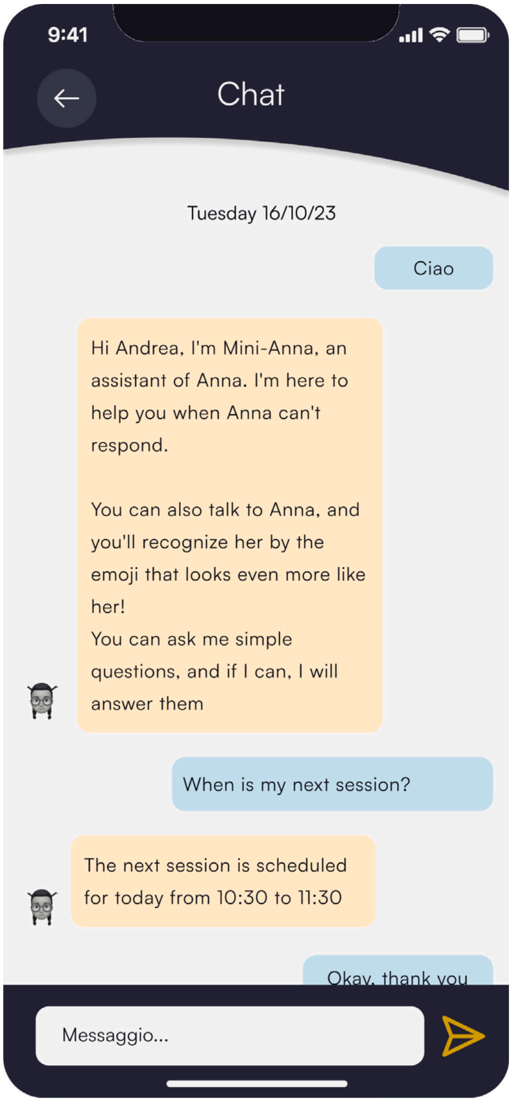


Fig. 7. Interface depicting the Chatbot.

colour) which garnered a completion rate of respectively 72,5% and 77,5%. In addition to the success rate score, the accuracy of participants' responses was evaluated by analysing the errors they made. Errors were identified as non-critical and given a value of 1 or critical and given a value of 0.

The occurrence of critical errors aligned with task completion rates, with Task 4 (Accessing the progress section and score monitoring) recording 2 critical errors (e.g., failing to access the Progress section) and 7 non-critical errors (e.g., having difficulties identifying the button in the navigation menu to access the Progress section or monitoring the current score) and Task 13 recording 3 critical errors (e.g., failing to access the colour customisation mode) and 3 non-critical errors (e.g., experiencing difficulties identifying the button to access the colour customisation mode or switching the eye colour). Additionally, non-critical errors were identified in Task 5 (e.g., having difficulties identifying the button to create a calendar entry), Task 6 and Task 7 (e.g., having difficulties placing the Lessons and Feelings Wheel in the Tasks section). However, the great majority of tasks were easily comprehended and completed by users.

The main error observed in Task 4 (Accessing the progress section and score monitoring) was the failure of participants to understand the Call To Action (CTA), or button in the footer, which would have directly navigated them to the Status page. This error demonstrates that, although the design is acceptable, there is still a need to modify the design, making the button clearer for the future version that will be used by larger samples of clinical users. Conversely, errors in Task 13 were primarily associated with difficulties in understanding the design of the colour customisation feature.

Task success rates for each task are presented in Table 1.

Table 1
The success rate for the tasks was calculated by assigning a score value of 0 = tasks not completed, 1 = tasks completed with difficulties, and 2 = tasks easily completed. The assigned scores were then summed and divided by the maximum possible score (40). This result was subsequently converted into a percentage.

#	Success rate	Tasks
Task 1	97.50%	Accessing and checking notifications
Task 2	95%	Accessing the chat and replying to the therapist
Task 3	97,50%	Accessing the Profile section and consulting personal data
Task 4	72.50%	Accessing the Progress section and score monitoring
Task 5	87.50%	Accessing the Calendar section and creating a calendar entry
Task 6	87.50%	Accessing the Lessons section
Task 7	80.00%	Compiling the Feelings Wheel and the related assignments
Task 8	95.00%	Compiling the PQ-16 survey
Task 9	97.50%	Compiling the food diary
Task 10	97,50%	Compiling the cognitive diary
Task 11	95.00%	Accessing the Profile section and the avatar customisation
Task 12	92.50%	Customising the bottoms and the emotion of the avatar
Task 13	77.50%	Accessing the colour customisation mode and switching the eye colour
Task 14	97.50%	Accessing the store and buying an item 29

3. Satisfaction. Satisfaction is defined as the mean value of the responses to the post-test administered SUS questionnaire questions, its total score ranges from 0 to 100, with higher values indicating higher satisfaction. Generally, a score exceeding 70 is regarded as good, whereas a score over 85 is deemed excellent (Ferreira et al., 2023). It can be represented by the following formula:

$$\text{Satisfaction} = \frac{\text{Question Value 1} + \text{Question Value 2}}{2} \tag{2}$$

Regarding the post-test administered SUS questionnaire, the average score is 74.37 (SD 3.88), which indicates that the application is perceived to be highly usable. The measured scores ranged from the lowest value being 67,5 (satisfying usability) and the highest score being 80 (good usability), demonstrating a consistent perceived usability in participants. However, the lowest score indicates the need to modify certain design elements of the application such as the button in the footer which leads the user to find its progress section and score monitoring.

4. Feedback. Following the testing sessions, participants provided valuable feedback on various aspects of the application. The feedback gathered from user interviews consistently highlighted the intuitive nature of the PEnguIN app, in fact out of the 20 participants, 17 (85 percent) provided positive feedback regarding the ease of navigation and user-friendly interface. Specific comments included mentions of the clear layout, logical flow of tasks, and straightforward instructions. Additionally, many participants commended the innovative inclusion of a TE system in mental health support, recognising its potential to enhance engagement and motivation during therapy. Their feedback suggested refining the implementation to integrate with therapy processes seamlessly and ensuring that token rewards offer tangible benefits, such as a new smartphone, a PlayStation, or hobby-related

Table 2
Comparisons between demographics and usability metrics.

User characteristics	Success rate (mean)	Error rate (mean)	Satisfaction (mean)	# of users
Gender				
Male	88,6%	2,2	74,3%	11
Female	89,9%	2,2	73,6%	9
Age				
18–23	90,7%	2	74,6%	12
23–27	90,6%	2,1	74,7%	8
Smartphone proficiency				
Moderate	85,36%	3	71,7%	10
Advanced	95,8%	1,1	77,5%	6
Extensive	96,4%	1	76,2%	4
Familiarity well-being applications				
Limited	84,12%	3,11	72,2%	9
Moderate	93,45%	1,8	73,7%	6
Extensive	97,85%	0,6	79%	5

products. Interestingly, while most participants expressed enthusiasm for the gamification features, two individuals appeared indifferent to this aspect. One stated that they never download any applications which use game-design features, and the other suggested that they would not find the rewarding system interesting. It is reasonable to believe that individuals who are not familiar with gamification features might benefit from a different motivational approach in therapy, focusing more on their progress and emphasising that gamification features are meant to support their personal growth rather than serving as an end in themselves. Participants also appreciated the creative use of metaphors, such as the symbolism of cherry blossom growth, to signify progress and personal development.

5. Demographics and usability metrics. The demographics analysis, coupled with objective usability metrics such as success rate, error rate, and satisfaction rate, provides additional insights into the user experience.

The data, shown in Table 2 indicate minimal variation in performance across different genders and age groups. The primary differences in performance were observed primarily between varying levels of experience with smartphones and well-being applications. As anticipated, individuals with *Extensive* experience with well-being applications achieved a success rate of 97.85%, whereas those with *Limited* experience attained a success rate of 84.12%. This suggests that even users with minimal prior exposure to such applications find the system highly usable. Additionally, variations in smartphone experience also yielded high success rates, with users possessing *Moderate* smartphone experience achieving an 85.5% success rate and those with *Extensive* smartphone experience reaching a nearly excellent 96.43%. These findings imply that, although technological experience enhances usability, the system is intuitively designed and remains effective for users with less experience, thereby demonstrating robust usability across diverse user demographics.

7. Conclusion and future work

This study marks the initial phase of a comprehensive effort to digitise the therapeutic process. It establishes a foundation for future research on digital health applications, particularly by providing a framework for evaluating usability and effectiveness within TE and gamification contexts.

Our research focused on the digitalisation of TE and gamification strategies for therapeutic use in the PEnguIN application, and it is perceived usability. While not intended to replace traditional instruments, PEnguIN serves as a valuable digital resource with the potential to advance future psychological assessments and treatments. Prioritising a user-friendly interface, we incorporated gamification elements to enhance engagement, therapeutic adherence, and lower dropout rates. Furthermore, PEnguIN extends digitisation beyond tools by incorporating treatment methodologies like a digitised TE system, making it a comprehensive digital solution for supporting psychological interventions in young adults.

These findings provide important insights into the perceived usability of the application, particularly its TE and gamified elements. The mixed-methods research approach, which combined qualitative and quantitative analyses, enriched our understanding of user interactions with the application, with the think-aloud method providing real-time insights into user experiences, revealing specific usability issues such as difficulties in locating certain features and understanding interface elements.

According to ISO 9241-11 (International Organization for Standardization, 2018a) establishing usability measures involves specifying intended goals, describing the context of use, and defining target values for effectiveness, efficiency, and satisfaction. We established effectiveness and satisfaction for PEnguIN. In fact, the high success rate and satisfaction scores across all groups affirm that the system is well-designed to accommodate a broad range of users.

In terms of satisfaction, the SUS results reveal a high perception of usability, with an average score of 74.37 (SD 3.88). Comparing this to recent studies (Hyzy et al., 2022), which establish a benchmark for satisfaction in mHealth applications at 68 (SD 12.5), it is evident that the application has strong usability foundations.

Overall, our application demonstrated strong performance across a range of effectiveness measures. However, in line with our dedication to ongoing improvement, we have found specific areas where further enhancements can be made. While the majority of tasks met our desired standards, a few stand out as candidates for refinement. Tasks 4 (Accessing the Progress section and score monitoring) and Task 13 (Accessing the colour customisation mode and switching the eye colour), in particular, showed success rates below 80%, signalling opportunities for improvement. Given the importance of these tasks in driving engagement with the gamification and TE system, addressing these areas presents a significant opportunity for improvement.

Based on the usability test findings, there is a generally positive impression of the application and its use. However, an improved design needs to address critical issues associated with tasks (4 and 13), particularly, the focus should be placed on refining aspects of the interface concerning Task 4 (Accessing the Progress section and monitoring the score), which records a success rate of 72.50%, as if constitute a central element in the context of the gamification and TE system. Regarding Task 4, the following improvements are proposed:

- All navigation buttons will be placed on the same level, ensuring uniformity and ease of access;
- Each navigation item will be accompanied by a recognisable icon and a descriptive label. For instance, the button leading to the "Progress" section could include a progress journey icon and be re-labelled as "Track Progress" or "My Progress" to enhance clarity and user understanding.

Similarly, for Task 13 the following improvements are recommended:

- The button leading to the colour customisation mode will be made more prominent through enhanced styling, such as increased size, high-contrast colours, and recognisable icons (e.g., a paint palette);
- The layout of the customisation interface could be optimised to separate colour selection tools from other options (e.g., clothing or emotions), reducing visual clutter and simplifying the user journey;
- Descriptive labels will be added to each feature within the customisation mode, including the "Eye Colour" option, ensuring users can immediately identify and use this functionality.

Following this, the next steps will involve conducting an extensive usability test to demonstrate the efficacy of the TE and gamification system over the long term. Moreover, trials involving the newly refined version of the application and individuals with clinical conditions (e.g., UHR, FEP) will follow, as the unique characteristics of our target population require thorough testing of our approach. In this regard, the present study served as a successful preliminary test of the app, identifying areas for improvement to ensure that the next version will be better suited for clinical users. Moreover, considering the continuum from non-clinical to clinical conditions, they can be regarded as analogous in terms of usability testing, as the app will also be used for preventive intervention in non-clinical populations and at-risk individuals without a diagnosis. Additionally, even among individuals with clinical conditions like UHR and FEP, various functioning levels can be observed, including areas of high functioning and no impairments. It is critical to emphasise that this study is extremely important for various stakeholders, including health professionals, educators, informal caregivers (e.g. parents), and anyone who interacts with outpatients. In fact, the application has received positive interest from institutions in the field of mental health, both in the national health system and private ones, as well as at conventions and from other important Italian psychotherapy societies, who have shown interest in participating in future experimentation in the clinical and research field. Therefore, we expect to collect invaluable data on the efficacy of our proposal in the near future.

This upcoming phase holds the potential to reveal insights into how our digital therapeutic approach can be practically implemented and its impact in real-world settings.

CRediT authorship contribution statement

Marco Cremaschi: Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Conceptualization. **Giulia Rosemary Avis:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **An Qi Zhao:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Elia Guarnieri:** Writing – original draft, Software, Methodology, Data curation. **Anna Panzeri:** Writing – review & editing, Investigation, Conceptualization. **Andrea Spoto:** Writing – review & editing, Supervision, Conceptualization.

Table A.3
Participants involved in the evaluation. *Experience*: participant experience with a smartphone. *Familiarity*: participant familiarity with well-being application for smartphone.

#	Age	Gender	Occupation	Experience	Familiarity
P1	23	Female	Student	Moderate	Limited
P2	20	Male	Student	Extensive	Limited
P3	25	Female	Worker	Extensive	Moderate
P4	22	Male	Student	Moderate	Limited
P5	19	Female	Worker	Moderate	Limited
P6	27	Male	Worker	Advanced	Moderate
P7	21	Female	Student	Moderate	Moderate
P8	24	Male	Unemployed	Advanced	Extensive
P9	18	Female	Student	Extensive	Extensive
P10	26	Female	Student	Moderate	Moderate
P11	23	Female	Worker	Advanced	Limited
P12	20	Male	Unemployed	Moderate	Limited
P13	19	Male	Worker	Extensive	Extensive
P14	27	Male	Student	Moderate	Limited
P15	22	Male	Worker	Advanced	Moderate
P16	26	Male	Student	Advanced	Extensive
P17	20	Female	Student	Moderate	Moderate
P18	24	Male	Student	Moderate	Limited
P19	21	Female	Student	Moderate	Limited
P20	27	Male	Student	Advanced	Extensive

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is partially funded by the "Ministry of Enterprises and Made in Italy" within the framework "Innovation Agreements" (Accordi per l'innovazione), project code F/310240/01-04/X56.

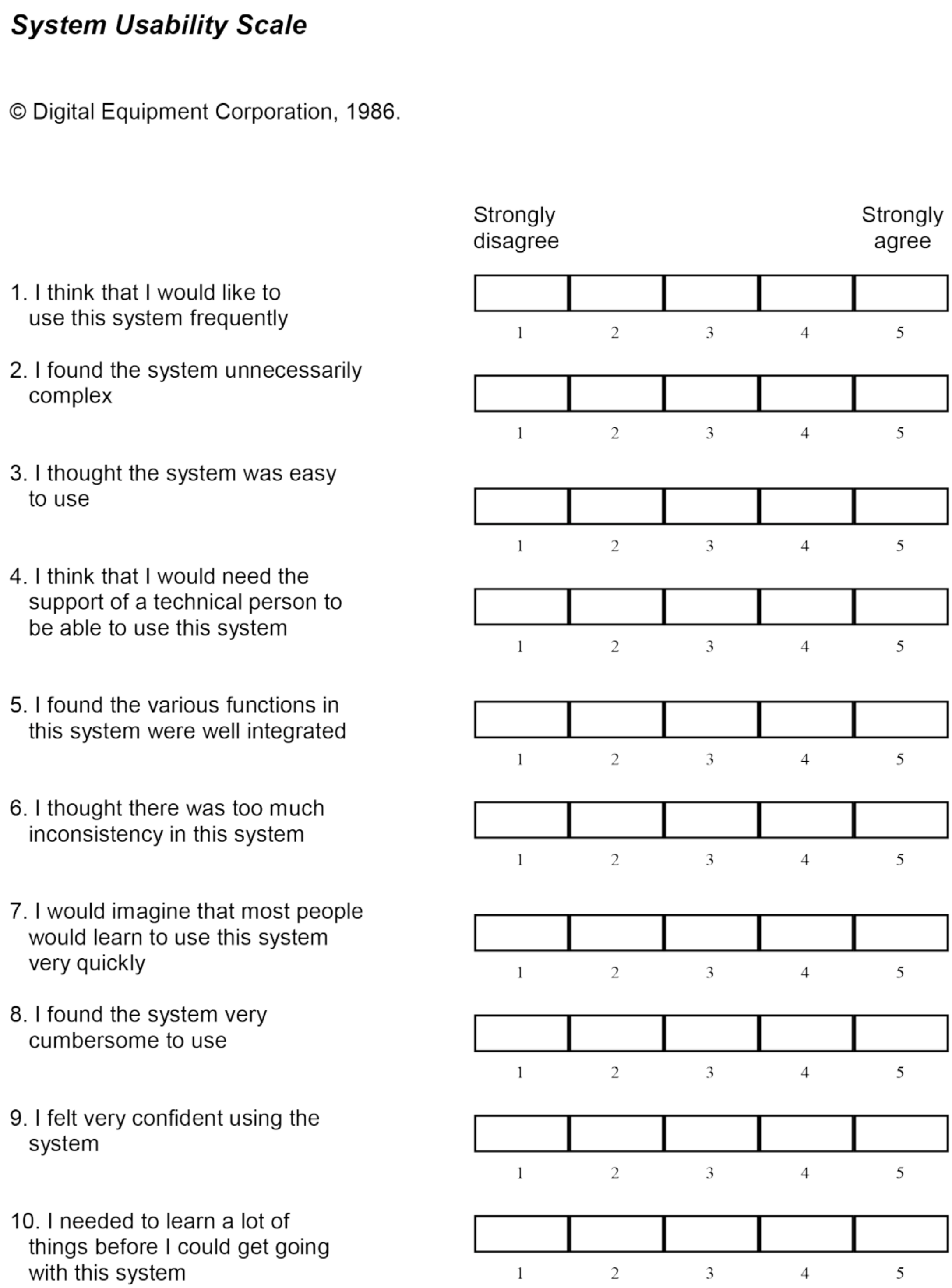
Appendix A. Participants table

See Table A.3.

Appendix B. Tasks

The tasks which were assigned to the users are the following:

- **Task 1**: Accessing and checking notifications
- **Task 2**: Accessing the chat and replying to the therapist
- **Task 3**: Accessing the Profile section and consulting personal data
- **Task 4**: Accessing the Progress section and score monitoring
- **Task 5**: Accessing the Calendar section and creating a calendar entry
- **Task 6**: Accessing the Lessons section
- **Task 7**: Compiling the Feelings Wheel and the related assignments
- **Task 8**: Compiling the PQ-16 survey
- **Task 9**: Compiling the food diary
- **Task 10**: Compiling the cognitive diary
- **Task 11**: Accessing the Profile section and the avatar customisation
- **Task 12**: Customising the bottoms and the emotion of the avatar
- **Task 13**: Accessing the colour customisation mode and switching the eye colour
- **Task 14**: Accessing the store and buying an item



System Usability Scale

© Digital Equipment Corporation, 1986.

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Fig. C.8. The SUS test from Brooke (1996).

Appendix C. SUS

See Fig. C.8.

Data availability

Data will be made available on request.

References

Addington, J., Farris, M., Devoe, D., & Metzak, P. (2020). Progression from being at-risk to psychosis: next steps. *npj Schizophrenia*, *6*(1), 27.

Amer, N. A., Shohieb, S. M., Eladrosy, W. M., Elbakry, H. M., & Abd Elrazek, S. M. (2023). Sokoon: A gamification-based cognitive behavioral therapy application–An application for depression, stress, and anxiety. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, *15*(1), 1–26.

Aparicio, A., Vela, F. L., González-Sánchez, J., & Isla-Montes, J.-L. (2012). Analysis and application of gamification.

Arciniegas, D. B. (2015). Psychosis. *Continuum (Minneap Minn)*, *21*, 715–736.

Bellack, A. S., Bennett, M. E., Gearon, J. S., Brown, C. H., & Yang, Y. (2006). A randomized clinical trial of a new behavioral treatment for drug abuse in people

with severe and persistent mental illness. *Archives of General Psychiatry*, *63*(4), 426–432.
Borrelli, B., & Ritterband, L. M. (2015). Special issue on ehealth and mhealth: Challenges and future directions for assessment, treatment, and dissemination. *Health Psychology*, *34*, 1205–1208.
Bouwmans, C., de Sonneville, C., Mulder, C. L., & Hakkaart-van Roijen, L. (2015). Employment and the associated impact on quality of life in people diagnosed with schizophrenia. *Neuropsychiatric Disease and Treatment*, 2125–2142.
Brooke, J. (1996). SUS: A 'quick and dirty'usability scale. Usability evaluation in industry. PW Jordan, b Thomas, BA Weerdmeester and AL Mcclelland.
Brown III, W., Yen, P.-Y., Rojas, M., & Schnall, R. (2013). Assessment of the health IT usability evaluation model (Health-ITUEM) for evaluating mobile health (mHealth) technology. *Journal of Biomedical Informatics*, *46*(6), 1080–1087.
Bucci, S., Barrowclough, C., Ainsworth, J., Morris, R., Berry, K., Machin, M., et al. (2015). Using mobile technology to deliver a cognitive behaviour therapy-informed intervention in early psychosis (Actissist): study protocol for a randomised controlled trial. *Trials*, *16*, 404.
Caon, M., Carrino, S., Condon, L., Ascolese, A., Facchinetti, S., Mazzola, M., et al. (2017). PEGASO companion: a mobile app to promote healthy lifestyles among adolescents. In *Wireless mobile communication and healthcare: 6th international conference, mobiHealth 2016, milan, Italy, November 14-16, 2016, proceedings 6* (pp. 53–61). Springer.
Chan, K. T. (2022). Emergence of the 'Digitalized Self' in the age of digitalization. *Computers in Human Behavior Reports*, *6*, Article 100191.
Chan, C. V., & Kaufman, D. R. (2011). A framework for characterizing ehealth literacy demands and barriers. *Journal of Medical Internet Research*, *13*(4), Article e1750.
Charters, E. (2010). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education : A Journal of Educational Research and Practice*, *12*, 68–82.
Csikszentmihalyi, M. (1975). Beyond boredom and anxiety (sıkılma ve anksiyetenin ötesi) San Francisco.
Csikszentmihalyi, M., & Csikzentmihaly, M. (1990). *Flow: The psychology of optimal experience*: *vol. 1990*, Harper & Row New York.
Dattani, S., Ritchie, H., & Roser, M. (2021). Mental health. *Our World in Data*.
Deo, N., Johnson, E., Kancharla, K., O'Horo, J. C., Kashyap, R., et al. (2020). Instant gratification as a method to promote physician practice guideline adherence: A systematic review. *Cureus*, *12*(7).
Deterding, S., Dixon, D., Khaled, R., & Nacke, L. E. (2011). *From game design elements to gamefulness: defining "gamification"* (pp. 9–15). MindTrek '11: ACM.
Doll, C., McLaughlin, T., & Barretto, A. (2013). The token economy: A recent review and evaluation. *International Journal of Basic and Applied Science*, *2*(1), 131–149.
Eccles, D. W., & Arsal, G. (2017). The think aloud method: what is it and how do I use it? *Qualitative Research in Sport, Exercise and Health*, *9*(4), 514–531.
Ferreira, J. M., Rodríguez, F. D., Santos, A., Dieste, O., Acuña, S. T., & Juristo, N. (2023). Impact of usability mechanisms: A family of experiments on efficiency, effectiveness and user satisfaction. *IEEE Transactions on Software Engineering*, *49*(1), 251–267.
Firth, J., & Torous, J. (2015). Smartphone apps for schizophrenia: A systematic review. *JMIR mHealth uHealth*, *3*(4), Article e102.
Fusar-Poli, P., Estradé, A., Stanghellini, G., Venables, J., Onwumere, J., Messas, G., et al. (2022). The lived experience of psychosis: a bottom-up review co-written by experts by experience and academics. *World Psychiatry*, *21*(2), 168–188.
Ghezzi, P. M., & Lewon, A. B. (2022). The token economy. In J. B. Leaf, J. H. Cihon, J. L. Ferguson, & M. J. Weiss (Eds.), *Handbook of applied behavior analysis interventions for autism: integrating research into practice* (pp. 497–511). Cham: Springer International Publishing.
Glowacki, K., Warner, G., & White, C. (2016). The use of a token economy for behaviour and symptom management in adult psychiatric inpatients: A critical review of the literature. *Journal of Psychiatric Intensive Care*, *12*(2), 119–127.
Glynn, S. M. (1990). Token economy approaches for psychiatric patients: Progress and pitfalls over 25 years. *Behavior Modification*, *14*(4), 383–407.
Gonzales, A. L., & Hancock, J. T. (2011). Mirror, mirror on my Facebook wall: Effects of exposure to Facebook on self-esteem. *Cyberpsychology, Behavior, and Social Networking*, *14*(1–2), 79–83.
Herrera, S. N., Sarac, C., Phili, A., Gorman, J., Martin, L., Lyallpuri, R., et al. (2023). Psychoeducation for individuals at clinical high risk for psychosis: a scoping review. *Schizophrenia Research*, *252*, 148–158.
Hjorthøj, C., Stürup, A. E., McGrath, J. J., & Nordentoft, M. (2017). Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. *The Lancet Psychiatry*, *4*(4), 295–301.
Horner, R., Hew, K. F., & Tan, C. Y. (2018). Comparing digital badges-and-points with classroom token systems: Effects on elementary school ESL students' classroom behavior and English learning. *Journal of Educational Technology & Society*, *21*(1), 137–151.
Hyzy, M., Bond, R., Mulvenna, M., Bai, L., Dix, A., Leigh, S., et al. (2022). System usability scale benchmarking for digital health apps: Meta-analysis. *JMIR Mhealth Uhealth*, *10*(8), Article e37290.
International Organization for Standardization (2006). *Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Common Industry Format (CIF) for usability test reports*: *Standard*, Geneva, CH: International Organization for Standardization.
International Organization for Standardization (2018a). *Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*: *Standard*, Geneva, CH: International Organization for Standardization.
International Organization for Standardization (2018b). ISO 9241-11:2018 Ergonomics of Human-System Interaction – Part 11: Usability: Definitions and Concepts.
Johansson, S., Gulliksen, J., & Lantz, A. (2015). Cognitive accessibility for mentally disabled persons. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, & M. Winckler (Eds.), *Human-computer interaction – INTERACT 2015* (pp. 418–435). Cham: Springer International Publishing.
Kazantzis, N., Deane, F. P., Ronan, K. R., & L'Abate, L. (2005). *Using homework assignments in cognitive behavior therapy*. Routledge.
Killikelly, C., He, Z., Reeder, C., Wykes, T., et al. (2017). Improving adherence to web-based and mobile technologies for people with psychosis: systematic review of new potential predictors of adherence. *JMIR mHealth and uHealth*, *5*(7), Article e7088.
Lancini, M. (2020). *Il ritiro sociale negli adolescenti: La solitudine di una generazione iperconnessa*. Raffaello Cortina Editore.
Lewis, J. R. (2018). The system usability scale: Past, present, and future. *International Journal of Human Computer Interaction*, *34*(7), 577–590.
Litvin, S., Saunders, R., Jefferies, P., Seely, H., Pössel, P., & Lüttke, S. (2023). The impact of a gamified Mobile mental health app (eQuoo) on resilience and mental health in a student population: large-scale randomized controlled trial. *JMIR Mental Health*, *10*, Article e47285.
McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin.
McGonigal, J. (2016). *SuperBetter: The power of living gamefully*. Penguin.
Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, *71*, 525–534.
Meneghelli, A., Alpi, A., Cascio, M., Häfner, H., Maurer, K., Preti, A., et al. (2013). Versione Italiana dell'Early Recognition Inventory for the retrospective assessment of the Onset of schizophrenia Checklist: affidabilità, validità e istruzioni per l'uso. *Journal of Psychopathology*, *19*, 1–2.
Moumane, K., Idri, A., & Abran, A. (2016). Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards. *SpringerPlus*, *5*, 1–15.
Nielsen, J. (2009). Discount usability: 20 years.
Nielsen, J., & Levy, J. (1994). Measuring usability: preference vs. performance. *Communications of the ACM*, *37*(4), 66–75.
Norman, C. D., & Skinner, H. A. (2006). Ehealth literacy: essential skills for consumer health in a networked world. *Journal of Medical Internet Research*, *8*(2), Article e506.
Ologundudu, O. M., Palaniyappan, L., Cipriano, L. E., Wijnen, B. F., Anderson, K. K., & Ali, S. (2023). Risk stratification for treating people at ultra-high risk for psychosis: A cost-effectiveness analysis. *Schizophrenia Research*, *261*, 225–233.
Rahill, K. M., & Sebrechts, M. M. (2021). Effects of Avatar player-similarity and player-construction on gaming performance. *Computers in Human Behavior Reports*, *4*, Article 100131.
Robacker, C. M., Rivera, C. J., & Warren, S. H. (2016). A token economy made easy through ClassDojo. *Intervention in School and Clinic*, *52*(1), 39–43.
Roldán-Ruiz, A. M., Merino-Godoy, M.-d.-l.-Á., Peregrín-Rubio, A., Yot-Dominguez, C., & da Costa, E. I. M. T. (2024). Assessing the hands-on usability of the healthy jeart app specifically tailored to Young users. *12*, In *Healthcare* (3), (p. 408). MDPI.
Stephenson, C., Moghimi, E., Gutierrez, G., Jagayat, J., Layzell, G., Patel, C., et al. (2023). User experiences of an online therapist-guided psychotherapy platform, OPTT: A cross-sectional study. *Internet Interventions*, *32*, Article 100623.
Takano, M., & Taka, F. (2022). Fancy avatar identification and behaviors in the virtual world: Preceding avatar customization and succeeding communication. *Computers in Human Behavior Reports*, *6*, Article 100176.
Taylor, T. (2002). Living digitally: Embodiment in virtual worlds.
Walker, E. F., Trotman, H. D., Goulding, S. M., Holtzman, C. W., Ryan, A. T., McDonald, A., et al. (2013). Developmental mechanisms in the prodrome to psychosis. *Development and Psychopathology*, *25*(4pt2), 1585–1600.
Weichbroth, P. (2024). Usability testing of mobile applications: A methodological framework. *Applied Sciences*, *14*(5), 1792.
Wenzel, A. (2017). Basic strategies of cognitive behavioral therapy. *Psychiatric Clinics of North America*, *40*(4), 597–609, Cognitive Behavioral Therapy for Anxiety and Depression.
Williamson, R. L., & McFadzen, C. (2020). Evaluating the impact of token economy methods on student on-task behavior within an inclusive Canadian classroom. *International Journal of Technology and Inclusive Education (Ijtie)*, *9*(1), 1531–1541.
Winnicott, D. (2005). *Playing and reality*. Routledge.
World Health Organization (2016a). *World mental health report: Transforming mental health for all*: *Report*.
World Health Organization (2016b). *Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth*: *Meeting report*.
Yee, N., & Bailenson, J. N. (2009). The difference between being and seeing: The relative contribution of self-perception and priming to behavioral changes via digital self-representation. *Media Psychology*, *12*(2), 195–209.
Zayeni, D., Raynaud, J.-P., & Revet, A. (2020). Therapeutic and preventive use of video games in child and adolescent psychiatry: a systematic review. *Frontiers in Psychiatry*, *11*, Article 487802.

Zhang, D., & Adipat, B. (2005). Challenges, methodologies, and issues in the usability testing of mobile applications. *International Journal of Human-Computer Interaction, 18*(3), 293–308.

Zhao, S., Sampson, S., Xia, J., & Jayaram, M. B. (2015). Psychoeducation (brief) for people with serious mental illness. *Cochrane Database of Systematic Reviews*, (4).

Zhou, L., Bao, J., Setiawan, I. M. A., Saptono, A., Parmanto, B., et al. (2019). The mHealth app usability questionnaire (MAUQ): development and validation study. *JMIR mHealth and uHealth, 7*(4), Article e11500.