Virtual Reality (2026) 30:24
https://doi.org/10.1007/s10055-025-01251-9

SI.: PROTOTYPING AND DEVELOPING REAL-WORLD APPLICATIONS FOR EXTENDED REALITY

VR-ACT: transforming clinical evidence of acceptance and commitment therapy through virtual reality for subthreshold depression

Bogoan Kim[1] · Dayoung Jeong[2] · Yoonseon Choi[3] · Hyungsook Kim[3] · Kyungsik Han[2,3]

Received: 29 March 2025 / Accepted: 16 October 2025

Abstract
Subthreshold depression (StD), recognized as a precursor to major depressive disorder, can lead to significant functional impairment and social burden without timely intervention, posing critical challenges for traditional therapeutic approaches in terms of accessibility and sustained engagement. Harnessing the power of digital intervention can be one possible solution, and in this paper, we present *VR-ACT*, a virtual reality (VR) system developed through interdisciplinary collaboration (e.g., HCI, mental health). The aim of *VR-ACT* is to transform the six core principles of acceptance and commitment therapy (ACT) into VR content and provide immersive experiences tailored for individuals with StD. The feasibility of *VR-ACT* was evaluated in a quasi-experimental study with 56 participants (28 StD group, 28 non-depressed group). We employed user experience measures to assess the feasibility and user engagement potential of the system and investigated five hypotheses by analyzing unique behavioral patterns in the StD group through sensor and log data. The study results demonstrated that *VR-ACT* is a feasible approach for delivering immersive mental health support and revealed distinct behavioral patterns in the StD group that can serve as potential clinical indicators. Our study contributes to the digital transformation of ACT by demonstrating the feasibility of *VR-ACT* and identifying digital phenotypes specific to individuals with StD. Building on these findings, we propose a digital transformation framework that emphasizes (1) mapping clinical evidence onto user interaction data and (2) fostering interdisciplinary collaboration led by experts in HCI and immersive system design to develop evidence-based digital interventions. We also outline key design implications for enhancing user engagement in immersive therapeutic settings.

Keywords Virtual reality · Acceptance and commitment therapy · Subthreshold depression · Digital transformation

✉ Kyungsik Han
kyungsikhan@hanyang.ac.kr

Bogoan Kim
bogoan@cbnu.ac.kr

Dayoung Jeong
dayoungjeong@hanyang.ac.kr

Yoonseon Choi
yoonseonchoi@hanyang.ac.kr

Hyungsook Kim
khsook12@hanyang.ac.kr

1 School of Information and Communication Engineering, Chungbuk National University, 1 Chungdae-ro, Seowon-gu, Cheongju 28644, Republic of Korea

2 Department of Artificial Intelligence, Hanyang University, 222 Wangsimni-ro, Seoul 04763, Republic of Korea

3 Department of Data Science, Hanyang University, 222 Wangsimni-ro, Seoul 04763, Republic of Korea

1 Introduction

Subthreshold depression (StD) refers to a condition in which individuals exhibit clinically significant depressive symptoms that do not meet the diagnostic criteria for major depressive disorder (MDD). First conceptualized by Judd et al. (1994), StD is now widely recognized as a precursor to MDD. It is characterized by the persistence of at least two depressive symptoms, such as low mood or loss of interest, for more than two weeks, accompanied by social impairment (Rodríguez et al. 2012). Over the past two decades, self-report depression scales have been used to objectively assess StD, with a widely adopted criterion defining StD as a Patient Health Questionnaire-9 (PHQ-9) score between 5 and 14 (mild to moderate depression) (Hao et al. 2023). This criterion has contributed to the consistency of StD research

Published online: 01 December 2025

across multiple studies (Wang et al. 2018; Rodríguez et al. 2012). The prevalence of StD has been steadily increasing, with recent meta-analyses estimating that approximately 11% of the adult population experiences StD (Zhang 2023), and rates increasing to as high as 47.8% during the COVID-19 pandemic (Xiong et al. 2020).

Early detection and intervention for StD are crucial not only for improving the quality of life of individuals, but also from a public health perspective (Hao et al. 2023; Xiong et al. 2020). StD can lead to significant social burden and functional impairment and is a major risk factor for the development of MDD (Cuijpers and Smit 2004). Individuals with StD experience reduced quality of life, impaired occupational functioning, and difficulties in social relationships (Zhang 2023; Hao et al. 2023). Moreover, StD is associated with increased suicide risk and long-term functional impairment, emphasizing the importance of early detection and intervention (Jeuring et al. 2016). However, individuals with StD often underestimate the severity of their symptoms, delay seeking treatment, or avoid seeking help due to the stigma surrounding mental health (Henderson et al. 2013), which can lead to worsening of symptoms and progression to more severe forms of depression (Cuijpers and Smit 2004).

Given the mild nature of StD symptoms, non-pharmacological interventions such as cognitive behavioral therapy (CBT), behavioral activation therapy (BAT), and Acceptance and Commitment Therapy (ACT) are often considered more appropriate than pharmacological treatments (Hao et al. 2023), whose risks may outweigh the benefits (Jiang et al. 2021). These interventions focus on providing individuals with coping skills and strategies to effectively manage symptoms without the side effects associated with medication (Farah et al. 2016; Gartlehner et al. 2017). However, traditional face-to-face psychotherapy poses spatial and temporal limitations, and individuals with StD may find it challenging to maintain consistent engagement in therapy due to practical difficulties, such as scheduling conflicts or transportation issues, and underestimation of the need for treatment (Solomon et al. 2001; Bertha and Balázs 2013; Sun et al. 2022).

Digital interventions have emerged as an important alternative to overcome these limitations (Moshe et al. 2021; Garrido et al. 2019; Linardon et al. 2019; Li et al. 2014; Kim et al. 2020). Digital tools such as Virtual Reality (VR), mobile apps provide easier access to treatment and can promote therapeutic engagement through personalized notifications and interactive features (Torous et al. 2021; Kammler-Sücker et al. 2023; Jin et al. 2025; Kim et al. 2024; Lee et al. 2020). Among these digital tools, VR has recently received considerable attention in recent years (Zhu 2022; Reese et al. 2022; Phelan et al. 2023; Spyridonis et al. 2024; Kim et al. 2024; Kim et al. 2024). Research has found that VR can augment psychotherapy by enabling individuals to experience challenging situations in a safe, simulated environment (Xu et al. 2021; Cieślik et al. 2023) and to practice emotional and cognitive coping strategies in a realistic and immersive manner (Kim et al. 2023; Dwivedi et al. 2022; Kim et al. 2022). Moreover, the potential of VR to capture and analyze user interactions during these therapeutic experiences has been recognized as a valuable asset for supporting mental health improvements (Torous et al. 2021). In particular, the integration of ACT to VR environments has been proposed as a suitable approach for individuals with StD (Juul et al. 2019). Unlike traditional symptom-focused treatments, ACT emphasizes the acceptance of difficult thoughts and emotions, while encouraging individuals to find meaning and value in those experiences (Hayes et al. 2006). Although several ACT-based mobile applications, such as ACT Companion (Harris 2022) and IntelliCare (Mohr et al. 2017), provide structured guidance for cognitive exercises and values clarification, they largely rely on user-initiated inputs, text-based selections, or static content delivery. As a result, these interaction modalities provide limited behavioral data granularity and lack embodied engagement, potentially constraining the experiential delivery of key ACT processes such as cognitive defusion and self-as-context.

VR-based systems can address these limitations by providing interactive and embodied environments that align closely with the experiential foundation of ACT. Such systems enable users to engage with therapeutic content in realistic yet controlled settings, allowing for safe exposure to emotionally challenging experiences. These affordances highlight the potential of VR to support experiential therapies (e.g., ACT). However, beyond selecting an appropriate delivery medium, it is equally important to reconsider how digital interventions are evaluated. Previous studies on digital mental health interventions have primarily assessed outcomes through pre- and post-intervention symptom measures. Recently, much attention has been paid to using data such as sensor signals from devices and log data from system use during the interaction process to understand user behavior patterns that can be quantitatively identified and characterized for clinical interpretation (Cho et al. 2023; Inc 2024). An important point here is to establish the relationship between interaction patterns and clinical evidence. This requires the design of content consistent with these relationships and the validation of hypotheses through experimentation. The process of designing digital content that includes the components derived from clinical evidence, formulating the hypotheses, collecting data that reflect these components, and conducting data analysis to test the hypotheses should be tightly integrated. Nonetheless, existing studies often lack a coherent link between these steps and have instead focused on clinical interpretation *after* data analysis. While this post-hoc interpretation approach is certainly valid, the results can vary significantly from study to study, and it would be difficult to present generalizable interpretations from a clinical perspective.

In this paper, we developed *VR-ACT* (Fig. 1), a system that effectively translates ACT into a VR environment through interdisciplinary collaboration between mental health professionals (MHPs), designers, and developers. This study focuses on demonstrating the feasibility and behavioral relevance of the system, rather than evaluating symptom reduction or long-term therapeutic effects. *VR-ACT* externalizes the six core principles of ACT—(1) *acceptance*, (2) *cognitive defusion*, (3) *being present*, (4) *self-as-context*, (5) *values*, and (6) *committed action*—over the course of five sessions. Detailed protocols for implementing each ACT principle in VR were created, reviewed, and iteratively refined. The *VR-ACT* system preserves the core principles of ACT while leveraging the strengths of VR to provide an immersive and interactive therapeutic experience that enables participants with StD to effectively engage with the ACT principles.

To evaluate the feasibility of the *VR-ACT* system, we pose the following research questions:

- *RQ1:* Can the core principles of ACT be effectively delivered in a VR environment?
- *RQ2:* How do the interaction patterns differ between the StD group and the non-depressed (ND) group?

To answer these questions, we conducted a user study consisting of five sessions with 28 participants in the StD group and 28 in the ND group, using a quasi-experimental control group design. By including a control group (the ND group), we aimed to identify the unique behavioral patterns of the StD group in the VR environment and analyze them in relation to the core ACT principles implemented based on clinical guidelines. In this way, specific responses or behavioral patterns exhibited by the StD group in the *VR-ACT* system can be interpreted as different outcomes that may occur when ACT principles are successfully applied. To further assess the feasibility and user engagement of *VR-ACT*, we evaluated the user experience using three scales—System Usability Scale (SUS), Igroup Presence Questionnaire (IPQ), and NASA Task Load Index (NASA-TLX). Based on the study results, we confirmed the feasibility of *VR-ACT*, as participants reported positive experiences across usability, presence, and workload measures, indicating a transformation of ACT principles into the VR environment. In addition, through hypothesis testing, we identified meaningful clinical indicators of StD by uncovering behavioral patterns and sensor data within the *VR-ACT* system. These results suggest that *VR-ACT* not only provides an engaging and practical digital implementation of ACT but also contributes to the digital phenotyping of StD.

Our research contributes to advancing the field of digital mental health interventions, specifically for StD, by leveraging the capabilities of VR technology in the context of ACT, as follows:

- We developed *VR-ACT*, focusing on the digital transformation of the core principles of *VR-ACT* into a clinical guideline-based VR environment, to enhance the accessibility and engagement with ACT principles for individuals with StD.
- We demonstrated the feasibility of *VR-ACT* through a user study and validated the practicality and acceptability of the system.
- We identified behavioral patterns of individuals with StD and derived meaningful clinical indicators based on log and sensor data within the VR environment.

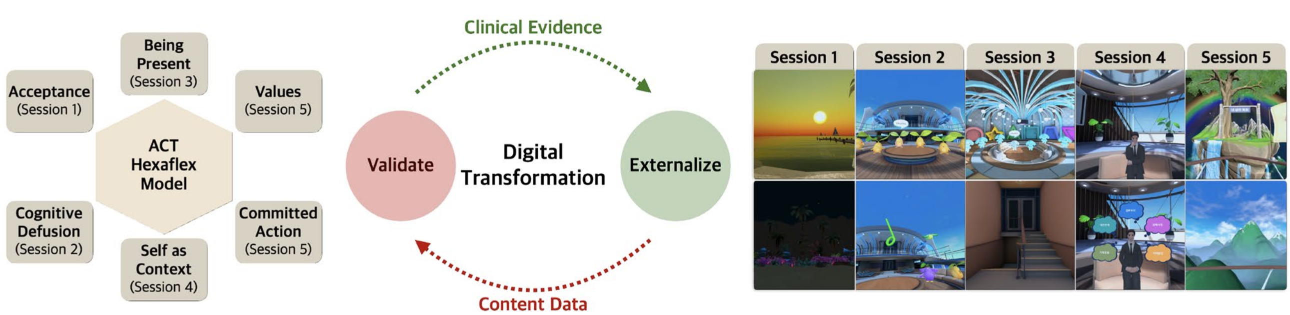

Fig. 1 Overview of VR-ACT demonstrating digital transformation that externalizes ACT principles into virtual environments, where participants mainly conduct given user tasks in the foreground (first row) while realistic background scenes (second row) enhance immersion across five sessions. The transformation process externalizes ACT principles based on clinical evidence, while content data collected from VR-ACT validates these core principles

2 Methods

In this section, we present the design and evaluation of *VR-ACT*, a VR-based system developed to translate ACT principles into immersive user experiences. We conducted a quantitative analysis of multimodal data, including psychological assessment measures (SUS, IPQ, NASA-TLX), *VR-ACT* user interaction logs, and sensor data (eye tracking, head movement) to evaluate the efficacy and feasibility of the *VR-ACT* system in delivering ACT interventions. This analysis compared the StD group with the ND group over five sessions.

2.1 VR-ACT system design

2.1.1 Theoretical foundation: ACT Hexaflex model

We designed *VR-ACT* based on the Hexaflex model (Bach and Moran 2008), a visual representation of the core principles of ACT in the form of a hexagonal diagram. This model illustrates six key processes of ACT, arranged around the central concept of psychological flexibility. The six key components are:

- *Acceptance:* The process of accepting negative thoughts and emotions without suppressing or avoiding them, and allowing them to be as they are.
- *Cognitive defusion:* The process of seeing thoughts as mere mental events and developing the ability to create distance from them.
- *Being present:* The process of cultivating the ability to focus on the present moment, free from concerns about the past or future.
- *Self-as-context:* The process of recognizing oneself as the observer of ever-changing thoughts and emotions, rather than identifying with them (e.g., noticing "I'm having a thought that I'm anxious" instead of "I'm anxious").
- *Values:* The process of clarifying meaningful life directions that are important to the individual.
- *Committed action:* The process of taking concrete actions in accordance with one's values.

Traditional ACT aims to increase psychological flexibility through these six processes. This approach teaches individuals to accept painful experiences rather than trying to eliminate them, and to live meaningfully within them. Although the processes in the Hexaflex model are interconnected, some should be structured to precede others. Specifically, *acceptance* and *cognitive defusion* should be addressed early as they form the basis for other processes. Conversely, *values* and *committed action* should be emphasized later as they build on the other processes.

Based on this model, we designed *VR-ACT* in five sessions. This structure provides sufficient time to experience each component while mitigating potential user discomfort (e.g., VR sickness). Sessions 1 through 4 sequentially reflect *acceptance*, *cognitive defusion*, *being present*, and *self-as-context*, while Session 5 addresses *values* and *committed action* together for an integrated application of ACT.

Sessions 1, 2, 4, and 5 took place in dynamic shared environments where users interacted with various virtual objects and agents—boarding a ship, receiving guidance from crew agents, creating and interacting with self-representing avatars, and navigating toward goals. In contrast, Session 3 uniquely allows users to enter and personalize their own space without interacting with virtual agents. This immersive and self-directed setting provides a psychologically safe context for practicing *being present* following *cognitive defusion*.

2.1.2 Determining user tasks in VR-ACT

The user tasks in *VR-ACT* are designed to allow users to experience and practice the six core components of ACT. These tasks are implemented through a narrative progression that begins on a beach, progresses to boarding a ship, and finally arrives at a destination that represents a user-selected goal. Each session includes key tasks, carefully selected based on clinical guidelines, which are important for evaluating the user's progress in adopting the principles of ACT (see Table 1). These tasks were collaboratively designed by MHPs, designers, and developers to effectively translate ACT principles into interactive VR experiences. The VR scenarios and key user interactions were refined through iterative co-design sessions, grounded in ACT manuals and therapeutic guidelines to ensure clinical relevance and fidelity to the foundational principles of the therapy.

The user tasks for each session are structured to allow the users to progressively experience and internalize the ACT principles. Early session tasks focus on having users recognize and accept their emotions and thoughts, while later sessions progress to tasks where users actively behave according to their values. The final session tasks involve actions that symbolize psychological growth, such as users taking the helm and steering toward their goals. All user tasks allowed unlimited interaction time, as the time taken to complete each task was considered a meaningful indicator for analyzing user engagement and task performance.

Fig. 2 illustrates the key scenes of user task performance in each session. By expressing the process of individual psychological growth as a symbolic journey, *VR-ACT* 's user tasks enhance engagement and motivation for treatment

Table 1 VR-ACT session structure with ACT Hexaflex components and user tasks

No.	Components	Main storylines	User Tasks
1	Acceptance	Participants arrive at a beach, initiating a journey of observing the present moment and exploring their emotions. Guided by narration, they observe their surroundings and experience recognizing and accepting their emotions	Meditate while observing the sea scenery Draw one's own constellation in the sky Listen to and advise an avatar facing difficulties **Collect emotion-labeled shells, practicing acceptance**
2	Cognitive defusion	Participants board a ship as passengers and receive guidance from the crew. They learn to recognize negative thoughts and distance themselves from them	Select negative thoughts that trouble oneself **Remove music bugs causing dissonance** Listen to music without dissonance Select their current emotion
3	Being present	Participants are guided to a secret space within the ship to focus on the present moment. They create a personal special place where they focus on current sensations and emotions	Choose the shape of the secret space **Decorate the space with various colors and meditate** Choose their current emotion
4	Self-as-context	Participants move to a regular cabin to experience an observer's perspective of themselves. They create an avatar similar to themselves, listen to its concerns, and offer advice, and see their thoughts and emotions from a new perspective	**Create an avatar resembling oneself** Select and listen to the avatar's concerns Choose and verbalize types of comfort to the avatar
5	Values, Committed action	Participants take control of the ship's helm and begin a journey towards their values. As the ship flies towards an aerial destination, they overcome various obstacles and practice actions committed to their values	**Select a flag with their important value written on it** Verbalize life goals related to the chosen value **Navigate towards life goals, overcoming obstacles**

The five sessions correspond to the components of the ACT Hexaflex model ("No." and "Components" refer to the session number and Hexaflex components). Each session includes the specific ACT component, a brief storyline description, and user tasks. Among these tasks, one key user task per session is highlighted in bold. These key tasks are designed to evaluate core ACT principles derived from the Hexaflex model based on clinical evidence

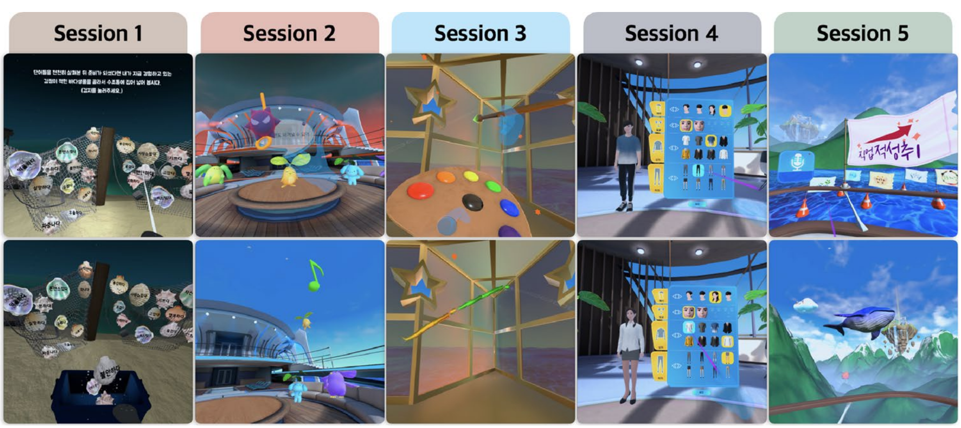

Fig. 2 Screenshots of the key user tasks across the five sessions, illustrating the main activities participants engaged in during each session. In Session 1, users collect emotion-labeled shells (top) and practice acceptance through meditation (bottom). In Session 2, they remove disruptive music bugs (top) and experience harmony (bottom). Session 3 involves decorating a personal space with various colors (top) and meditating (bottom). In Session 4, users create avatars that resemble themselves (top: male; bottom: female). In Session 5, participants select life goals (top) and navigate obstacles to pursue them (bottom)

while providing opportunities to learn and practice core ACT principles experientially. The inclusion of clinically significant key tasks ensures a robust evaluation of user progress throughout the *VR-ACT* experience.

2.1.3 Building a branching narrative for VR-ACT

To effectively implement *VR-ACT*, we adopted a branching narrative structure (Riedl and Young 2006). Unlike traditional linear narratives, this approach provides flexibility within a fixed framework, offering various pathways based on user choice while maintaining the overall treatment goals. In each *VR-ACT* session, users can select ACT practice activities that match their experience and preferences. This design is consistent with ACT's emphasis on personal values and context-aware interventions by respecting individual preferences and promoting mindful, values-based actions (Hayes et al. 2006). As a result, users can derive personal meaning from the *VR-ACT* experience, fostering stronger connections with the virtual environment and increasing motivation for continued engagement. The branching structure not only supports user autonomy, but also enables the collection of diverse interaction patterns, providing a valuable foundation for analyzing how individual choices relate to the practice of ACT principles.

2.1.4 Enhancing user immersion and engagement

VR-ACT incorporates a range of auditory elements to deepen user immersion and foster emotional and cognitive engagement throughout the therapeutic experience. Each session is designed to respond dynamically to user interaction, guiding cognitive and emotional responses through auditory stimulation. Specifically, in Session 2, background music changes based on the negative thoughts selected by the user. In Session 3, the rhythm of the music adjusts in real time to match the user's pace while decorating their selected secret spaces. Across all sessions, spoken voiceovers and on-screen prompts to support user comprehension and task flow. Each session concludes with opportunities for users to verbally express their emotions.

2.2 Research hypotheses

We proposed five hypotheses, associated with each session, to evaluate the feasibility clinical relevance of the *VR-ACT* system. As mentioned in Sections 2.1.1 and 2.1.2, each session was designed according to ACT principles and clinical guidelines, with clinically significant user tasks defined for sessions that emphasize these principles. By focusing on these externalized key user tasks, we developed hypotheses to assess different behavioral patterns between the StD and the ND groups in each session. Through this, we aim to examine whether *VR-ACT* successfully reflects ACT principles in both groups, while identifying significant clinical indicators unique to the StD group.

- *H1:* The StD group will exhibit differences in recognized and accepted emotions compared to the ND group.
- *H2:* The StD group will have greater difficulty in alleviating emotional burden during the cognitive defusion process compared to the ND group.
- *H3:* The StD group will show comparable levels of psychological engagement with the present moment compared to the ND group.
- *H4:* The StD group will show greater difficulty in self-representation tasks compared to the ND group.
- *H5:* The StD group will have difficulty clarifying their values and smoothly engaging in committed actions in Session 5 compared to the ND group.

All five hypotheses were based on core ACT principles derived from the Hexaflex model and established clinical evidence. Each hypothesis corresponds to one or two key user tasks, which were designed by translating the six ACT principles into interactive experiences. While each of H1, H2, H3, and H4 targeted a single ACT principle and key user task, H5 addressed two principles and was therefore associated with two key user tasks. These tasks facilitated a structured assessment of user engagement with ACT concepts in the VR environment.

2.3 Recruitment

To minimize the potential impact of VR technology as an exogenous variable influencing experimental results, we restricted the age of participants to between 18 and 40 years to ensure a more accurate assessment of the feasibility of *VR-ACT*. Participants were recruited through student community websites, flyers, and outreach via the mental health center at the authors' institution. All participants provided written informed consent prior to participation.

Initially, we recruited 36 young adults with subthreshold depression ($Mean = 27.3, SD = 4.2$) and 32 non-depressed young adults ($Mean = 26.8, SD = 3.9$) in the same age range, who were screened according to the following criteria: (1) young adults aged between 18 and 40, (2) able to visit our facility, (3) a score of 85 or higher on the Northstar Digital Literacy Assessment (NDLA), which evaluates basic digital skills (e.g., using document software), with a score of 85 representing a passing level of functional digital literacy according to the official NDLA guideline, and (4) a score on the PHQ-9, with subthreshold depression scored as five or higher and non-depressed

Table 2 Demographic information of participants in the StD and the ND groups, including age, gender, frequency of prior VR experience, and PHQ-9 score levels for depression severity (none-minimal: 0–4, mild: 5–9, moderate: 10–14, moderately severe: 15–19)

Characteristics	StD ($n = 28$)	ND ($n = 28$)
Age		
Years, *Mean* ± *SD*	23.43 ± 4.40	25.86 ± 4.35
Gender		
Female (%)	19 (67.86)	16 (57.14)
Male (%)	9 (32.14)	12 (42.86)
VR experience		
0 times (%)	7 (25.00)	5 (17.86)
1–2 times (%)	15 (53.57)	14 (50.00)
3–5 times (%)	4 (14.29)	7 (25.00)
More than 5 (%)	2 (7.14)	2 (7.14)
PHQ-9 score		
None-minimal (%)	–	28 (100)
Mild (%)	10 (35.71)	–
Moderate (%)	15 (53.58)	–
Moderately severe (%)	3 (10.71)	–

scored as below five. Although 5–14 is commonly used to indicate StD, three participants with scores of 15–16 were also included based on clinical judgment during the in-person screening, as their symptoms aligned with subthreshold characteristics. Participants first completed an online prescreening questionnaire and were rescreened in person at their first visit prior to final enrollment.

Exclusion criteria were applied as follows: (1) participants who did not meet the inclusion criteria ($n = 2$), (2) those with comorbid psychiatric conditions (e.g., bipolar disorder, schizophrenia) that impaired decision-making capacity, which required users to complete interactive tasks without researcher intervention ($n = 2$), (3) those who refused to participate in the experiment ($n = 4$), and (4) those who had technical issues that resulted in the partial data loss ($n = 4$). After applying these exclusion criteria, the final sample consisted of 28 participants in the StD group and 28 in the ND group. Table 2 presents demographic information, including PHQ-9 scores, gender, age, and prior VR experience of the participants.

2.4 Data collection

2.4.1 Measures of user experience and feasibility

We used three measures to evaluate the user experience and feasibility of *VR-ACT*. The IGroup Presence Questionnaire (IPQ) and the NASA Task Load Index (NASA-TLX) were measured immediately after each *VR-ACT* session, while the System Usability Scale (SUS) was evaluated once at the end of the entire program. The SUS consists of 10 items, with scores below 68 indicating inadequate usability, 68–84 indicating adequate usability, and 85 or above indicating excellent usability. The IPQ consists of 14 items, including one item measuring general presence and three sub-scales evaluating spatial presence, involvement, and realness. Each item is measured on a 7-point Likert scale, providing a comprehensive assessment of the user's presence in the virtual environment. The NASA-TLX measures the user's cognitive load during *VR-ACT* activities by evaluating six sub-domains: mental demand, physical demand, temporal demand, performance, effort, and frustration, using a 7-point Likert scale. Furthermore, we conducted session-specific questions to evaluate the application of the core principles of ACT in each session (Table 3). These questions were designed to assess whether the key ACT concepts were effectively implemented and understood by participants in the context of each *VR-ACT* session.

Table 3 Session-specific questions for evaluating ACT principles ("No." refers to the session number)

No.	ACT principle	Evaluation question
1	Acceptance	Did Session 1 help you accept negative emotions?
2	Cognitive defusion	Did Session 2 help you detach from your thoughts?
3	Being present	Did Session 3 help you focus on the present moment?
4	Self-as-context	Did Session 4 help you objectively observe yourself?
5	Values, committed action	Did Session 5 help you identify and act on your core values?

2.4.2 Sensor and log data

We unobtrusively collected real-time sensor data and user interaction logs during VR-ACT sessions using Unity3D Engine. Gaze data recorded at 90 Hz via Meta Quest Pro's eye-tracking technology, while head movement data was collected using built-in motion tracking sensors (e.g., gyroscope, accelerometer, magnetometer). User interactions with objects were recorded using a laser pointer; when an object was touched, the system recorded the object identifier and the corresponding touch signal, enabling identification of the user's region of interest (ROI). We also recorded user responses and actions during specific tasks in each *VR-ACT* session, including task choices and time spent on each task. This setup allowed for high-resolution tracking of user attention and movement within the virtual environment, providing a rich quantitative basis for analyzing user behavior.

2.5 Study procedure

We conducted a five-session study with 28 participants in the StD group and 28 in the ND group. All five sessions were completed during a single visit. The study procedure consisted of (1) a 45-min introductory phase, (2) a *VR-ACT* experience phase ($Mean = 58.34\,\text{min}, SD = 5.12\,\text{min}$), and (3) a 20-min review phase. During the introductory phase, researchers explained the overall study purpose and procedures. Participants were also informed about data types and data management procedures. They completed a pre-study questionnaire covering NDLA, prior VR experience, and basic demographic information. A brief VR interface tutorial was provided to familiarize participants with essential interface operations in the VR environment (e.g., button pressing, object selection). Throughout the *VR-ACT* experience phase, participants wore an HMD over five sessions. Each session followed a fixed order based on the flow of the hexaflex model. All task instructions were delivered through spoken and on-screen prompts, which minimized researcher intervention during the sessions and allowed participants to independently explore their emotions. After each session, they completed questionnaires including the IPQ and NASA-TLX to evaluate immersion and cognitive load. Session-specific questions were also presented to assess the proper application of ACT's core principles in each session (See Table 3). Participants were given sufficient rest if they reported physical discomfort at the end of each session. In the review phase, participants completed the SUS questionnaire to gather comprehensive feedback on the overall *VR-ACT* experience. This study was approved by the Institutional Review Board (IRB) of the authors' affiliated institution, and prior consent was obtained from each participant. To protect participants' privacy, all personally identifiable information was removed from the collected data, and code names were assigned for anonymization before analysis.

2.6 Data analysis

To evaluate the feasibility of *VR-ACT* and investigate unique behavioral patterns in participants with StD, we conducted descriptive statistics and statistical tests on post-evaluation results for several measures, including SUS, IPQ, NASA-TLX, ROI-based level of attention, and session-by-session log data. Our analysis aimed to identify significant differences between the StD and ND groups. To control for potential confounding variables, we included gender, age, and prior VR experience as covariates in the statistical model. Gender and age were selected based on their frequent inclusion in digital intervention studies, where they are known to influence user experience and attentional responses (Carl et al. 2019; Kober and Neuper 2013). Prior VR experience was incorporated based on clinical recommendations, as varying levels of familiarity with immersive environments may affect task engagement, even among relatively digitally literate young adults who completed the interface tutorial.

We used Analysis of Covariance (ANCOVA) to adjust for these covariates and ensure robust comparisons between groups. Before running ANCOVA, we confirmed all necessary conditions, including independence, linearity, homoscedasticity, normality of covariates, and the absence of interaction effects between covariates and treatments. The Shapiro-Wilk test was used to assess the normality of covariates and residuals, residual plots were examined to evaluate linearity, and the Levene's test was applied for homoscedasticity. For the ANCOVA results, we used partial eta squared (η_p^2) to estimate the effect sizes, with values around 0.01 interpreted as small, 0.06 as medium, and 0.14 as large effects.

In Session 3, unlike other sessions in which task completion time served as a meaningful performance indicator, the focus was on psychological engagement with the present moment. Participants were asked to perform immersive drawing and observation tasks in a private setting without interaction with virtual agents or characters. Saccade frequency can serve as an indicator of visual attentional stability and cognitive processing efficiency; frequent eye movements may imply increased attentional shifts during task performance, which are often associated with cognitive distraction or emotional instability (Rayner 1998; Gibb et al. 2016). Therefore, we employed saccade-based attention analysis to assess concentration levels. Specifically, we calculated saccades from eye and head movement data (gaze position *x*, *y*, *z*, and head rotation *roll*, *pitch*, *yaw*) collected from the HMD to evaluate whether participants were focused on the present moment (Equations 1-4). Based on prior research indicating that frequent saccades may suggest dispersed attention (Gu et al. 2024; Noyes et al. 2023), we defined saccades as occurring when eye movement velocity exceeded a set threshold while changes in head rotation remained below this threshold. In addition, we used Pearson correlation analyses to examine the relationship between avatar customization time and the total time spent in Session 4 for the StD and ND groups.

$$\Delta\text{position} = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2} \tag{1}$$

$$\text{velocity} = \frac{\Delta\text{position}}{\Delta t} \tag{2}$$

where $\Delta x = x_{i+1} - x_i$, $\Delta y = y_{i+1} - y_i$, and $\Delta z = z_{i+1} - z_i$. i refers to the value at the previous time point, and $i+1$ represents the value at the next time point. Δt is the time interval and is calculated as $t_{i+1} - t_i$.

$$\Delta\text{rotation} = \sqrt{(\Delta \mathrm{r}_x)^2 + (\Delta \mathrm{r}_y)^2 + (\Delta \mathrm{r}_z)^2} \tag{3}$$

where $\Delta r_x = r_{x_{i+1}} - r_{x_i}$, $\Delta r_y = r_{y_{i+1}} - r_{y_i}$, and $\Delta r_z = r_{z_{i+1}} - r_{z_i}$. The changes in head rotation were calculated for each rotational coordinate (*roll*, *pitch*, *yaw*) and then integrated into a three-dimensional rotational change.

$$\text{saccade} = \begin{cases} \text{True,} & \text{if velocity} > v_{\text{threshold}} \text{ and } \Delta\text{rotation} < r_{\text{threshold}} \\ \text{False,} & \text{otherwise} \end{cases} \tag{4}$$

where $v_{threshold}$ is the velocity threshold and $r_{threshold}$ is the rotational change threshold. Each threshold was set to 10, based on approximately 10% of the variation between the minimum and maximum values in the collected data.

3 Results

3.1 Evaluation of ACT digital transformation (RQ1)

We analyzed the survey results of IPQ, NASA-TLX, and session-specific questions (Table 3) to evaluate the user perception of ACT. SUS was collected once, completing all sessions, to assess the overall usability of the *VR-ACT* experience. Table 4 presents the mean and SD values for IPQ, NASA-TLX, and session-specific questions across sessions for both the StD and ND groups.

First, the SUS scores for the StD and ND groups were 81.57 (*SD* = 12.79) and 83.64 (*SD* = 8.02), respectively. Both were significantly higher than the general average of 68. There was no statistically significant difference between the two groups ($F(1, 50) = 0.19$, $p = 0.67$, $\eta_p^2 = 0.003$).

Second, in the IPQ, both the StD and ND groups showed similar scores in the general item. In Session 1, the scores were 5.36 (*SD* = 0.87) and 5.25 (*SD* = 1.29) for the StD and ND groups, respectively, and in Session 5, they were 6.04 (*SD* = 0.74) and 5.57 (*SD* = 1.10), respectively. There were no statistically significant differences between the two groups in Spatial Presence, Involvement, and Realism. These results suggest that depressive symptoms did not significantly influence participants' subjective sense of presence or immersion in the VR environment.

Third, some subscales of the NASA-TLX showed differences between the groups. For Mental Demand, the StD group reported higher cognitive workload in Session 2 with a score of 2.18 (*SD* = 1.52), compared to the ND group, which scored 1.50 (*SD* = 1.07). The analysis yielded $F(1, 50) = 4.90$, $p = 0.03$, $\eta_p^2 = 0.09$. For Temporal Demand, the

Table 4 Results of the three measures—IPQ, NASA-TLX, and session-specific questions—used to evaluate the digital transformation of ACT across each session

Measures	Session 1		Session 2		Session 3		Session 4		Session 5	
	StD	ND	StD	ND	StD	ND	StD	ND	StD	ND
IPQ										
General item	5.36 (0.87)	5.25 (1.29)	5.28 (1.27)	5.39 (1.25)	5.79 (1.10)	5.21 (1.47)	5.53 (1.20)	5.29 (1.49)	6.04 (0.74)	5.57 (1.10)
Spatial presence	5.25 (0.81)	5.40 (0.97)	5.31 (0.92)	5.26 (0.94)	5.40 (0.90)	5.13 (1.15)	5.15 (1.09)	5.37 (0.94)	5.69 (0.91)	5.50 (1.07)
Involvement	4.77 (1.34)	4.84 (1.51)	4.63 (1.58)	5.07 (1.50)	5.01 (1.49)	5.14 (1.41)	5.11 (1.48)	5.56 (1.40)	5.38 (1.26)	5.46 (1.33)
Realism	3.62 (1.12)	3.56 (1.24)	3.68 (1.29)	3.52 (1.33)	3.96 (1.38)	3.38 (1.37)	4.01 (1.55)	4.08 (1.56)	4.26 (1.66)	3.44 (1.52)
NASA-TLX										
Mental demand	1.86 (1.24)	1.54 (0.96)	*2.18 (1.52)*	*1.50 (1.07)*	2.29 (1.62)	1.61 (1.13)	1.64 (0.95)	1.36 (0.99)	1.71 (1.38)	1.32 (0.98)
Physical demand	1.50 (0.92)	1.50 (1.14)	1.68 (1.06)	1.39 (0.99)	2.04 (1.57)	2.00 (1.70)	1.36 (0.62)	1.39 (1.17)	1.71 (1.21)	1.46 (1.35)
Temporal demand	2.39 (1.57)	1.93 (1.39)	2.32 (1.66)	1.64 (0.99)	2.64 (1.85)	2.14 (1.56)	*2.25 (1.65)*	*1.36 (0.62)*	1.61 (0.99)	1.68 (1.19)
Effort	2.36 (2.06)	2.32 (2.00)	2.21 (1.79)	1.92 (1.88)	2.39 (1.66)	3.21 (2.10)	1.75 (1.21)	1.96 (1.86)	1.57 (1.23)	1.50 (1.23)
Performance	2.25 (1.62)	2.46 (2.01)	1.93 (1.44)	2.14 (2.01)	1.68 (1.31)	2.11 (1.85)	1.64 (1.16)	1.86 (1.58)	1.36 (0.56)	1.86 (1.88)
Frustration	1.75 (1.40)	1.32 (0.48)	***1.96 (1.50)***	***1.21 (0.42)***	1.57 (1.14)	1.29 (0.98)	*2.32 (1.59)*	*1.43 (0.88)*	1.57 (1.14)	1.14 (0.36)
Session-specific questions										
Overall score	4.14 (0.52)	4.07 (0.60)	***3.50 (0.70)***	***4.20 (0.68)***	3.68 (0.98)	3.82 (1.02)	*3.71 (1.08)*	*4.25 (0.97)*	4.39 (0.69)	4.43 (0.63)

Each measure is denoted by its abbreviation in the table. Values are presented as "*Mean* (*SD*)," and mean differences between groups are indicated in bold (italic for p <0.05 and bold and italic for p <0.01)

StD group reported a higher sense of time pressure in Session 4 ($F(1, 50) = 6.01$, $p = 0.02$, $\eta_p^2 = 0.11$), with the score of 2.25 (SD = 1.65), compared to the ND group, which reported a score of 1.36 (SD = 0.62), suggesting that the StD group felt more time pressure during the user task in this session.

Regarding Frustration, the StD group experienced generally higher levels than the ND group. Significant differences were observed in Session 2 ($F(1, 50) = 7.18$, $p < 0.01$, $\eta_p^2 = 0.12$) and Session 4 ($F(1, 50) = 6.61$, $p = 0.01$, $\eta_p^2 = 0.11$), indicating that the StD group may have experienced greater emotional burden when disengaging from intrusive thoughts and presenting themselves objectively during key tasks in these sessions.

Lastly, we analyzed the scores of the session-specific questions mentioned in Table 3 to evaluate the core ACT principles. In Sessions 1, 3, and 5, there were no significant differences between the two groups, indicating that both groups generally found these sessions helpful and engaging. Notably, statistically significant differences were observed in Sessions 2 and 4. In Session 2, the StD group scored 3.50 (*SD* = 0.70) while the ND group scored 4.20 (*SD* = 0.68) ($F(1, 50) = 13.58$, $p < 0.01$, $\eta_p^2 = 0.20$), and in Session 4, the StD group scored 3.71 (*SD* = 1.08) while the ND group scored 4.25 (*SD* = 0.97) ($F(1, 50) = 4.53$, $p < 0.05$, $\eta_p^2 = 0.08$). While both scores were within a positive range, the relative differences suggest that participants with StD may have experienced more difficulty disengaging from intrusive thoughts in Session 2 and adopting an observer perspective in Session 4. These findings are consistent with the Frustration subscale results, suggesting that cognitive defusion and self-observation were particularly demanding tasks for participants with StD. This aligns with ACT theory and clinical studies indicating that individuals with StD struggle to detach from intrusive thoughts and maintain an observer perspective (Hayes et al. 2006; Juul et al. 2019).

3.2 Hypothesis testing: distinct behavioral patterns of the StD group (RQ2)

3.2.1 Emotion recognition differences (H1)

We examined the differences in emotion recognition and acceptance between the StD and ND groups in Session 1. As shown in Fig. 3a, while ANCOVA analysis did not reveal significant differences in the number of positive emotions selected (positive emotion shell selection) between the two

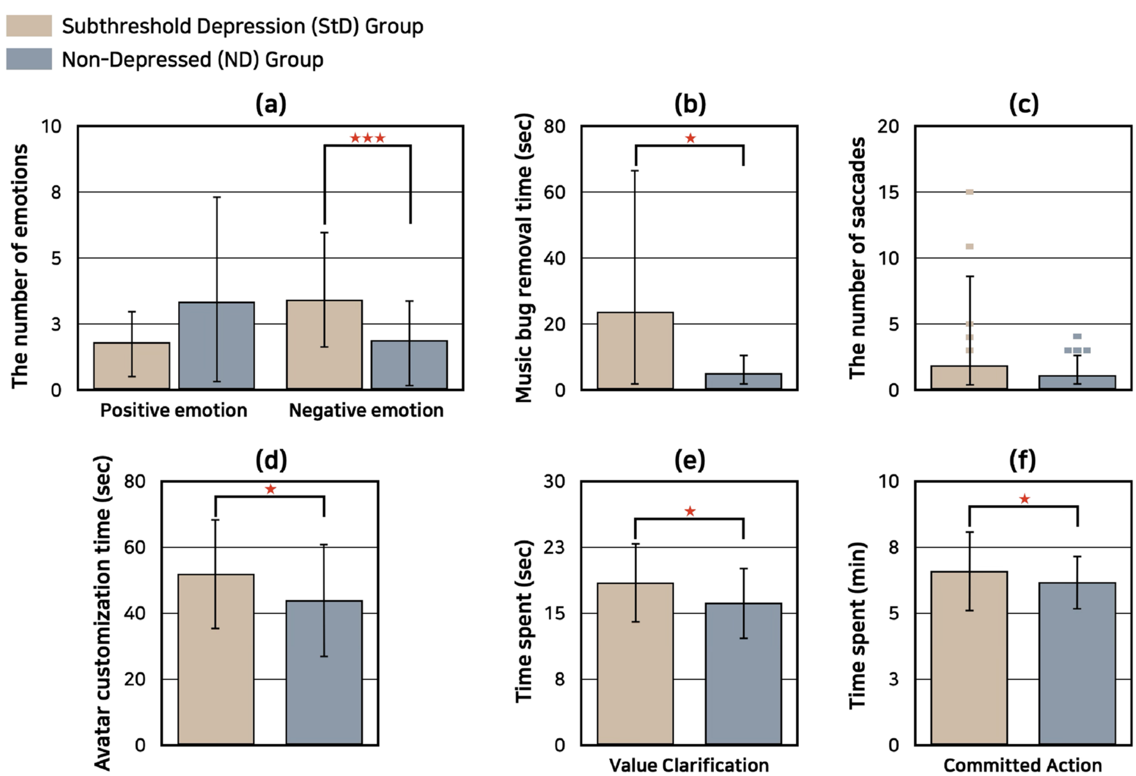

Fig. 3 Results for each hypothesis (H1–H5). **a** represents the number of positive and negative emotions for H1, **b**)represents the time taken to remove the music bugs for H2, **c** represents the number of saccades for H3; square-shaped dots indicate outliers, **d** represents the time spent on avatar customization for H4, and **e**, **f** represent the time spent on value clarification and committed action tasks, respectively, for H5. Error bars indicate 95% confidence intervals for each measure. Red stars indicate statistical significance (*$p < 0.05$, ***$p < 0.001$)

groups ($F(1,50)=2.79, p=0.10, \eta_p^2=0.05$), there was a statistically significant difference for negative emotions, with the StD group selecting more negative emotions than the ND group ($F(1,50)=15.16, p$ <0.001, $\eta_p^2=0.23$). This lack of significance in the selection of positive emotions is likely due to considerable individual variation in the ND group's responses (the StD group: $Mean=1.82, SD=1.36$; the ND group: $Mean=3.34, SD=3.95$), reducing statistical power. Specifically, the StD group selected an average of 3.43 negative emotions ($SD=2.40$) out of 25, whereas the ND group selected 1.89 ($SD=1.55$). This suggests that the StD group is more sensitive to recognizing and responding to negative emotions than the ND group.

3.2.2 Cognitive defusion difficulty (H2)

In Session 2, we assessed whether participants had difficulties relieving emotional burdens during the cognitive defusion process. The results showed that the StD and ND groups took an average of 23.75 s ($SD=54.06$) and 5.07 s ($SD=5.83$), respectively. The removal of the "music bug" ($F(1,50)=4.55, p=0.038, \eta_p^2=0.08$) showed a statistically significant difference (see Fig. 3b). This suggests that the StD group required more time to reduce emotional stress during the process.

3.2.3 Degree of present-moment retention (H3)

In Session 3, we evaluated whether participants made an effort to focus on the present moment by analyzing their degree of present-moment retention during the pattern drawing and observation activities.

The analysis revealed no statistically significant difference in the degree of present-moment retention between the two groups ($F(1,50)=1.89, p=0.17, \eta_p^2=0.037$). Both groups showed similar levels of engagement during the pattern drawing and observation tasks, indicating that the engagement of the StD group in tasks requiring present-moment focus was comparable to that of the ND group. The average number of saccades was slightly higher in the StD group ($Mean=1.86, SD=3.45$) compared to the ND group ($Mean=1.11, SD=1.20$), as illustrated in Fig. 3c).

3.2.4 Self-as-context delays (H4)

In Session 4, we examined whether participants had difficulty objectively recognizing their thoughts or emotions. As shown in Fig. 3d, the time taken to create an avatar showed a statistically significant difference between the two groups ($F(1,50)=4.76, p=0.034, \eta_p^2=0.08$). On average, the StD group spent 52.0 s ($SD=16.50$) on avatar customization, whereas the ND group spent 44.0 s ($SD=17.0$). This finding suggests that participants with StD may experience greater cognitive or emotional difficulty during self-as-context tasks that involve abstract self-representation. Furthermore, the correlation between the total time spent in Session 4 and the time spent on avatar customization was significant for both the StD group ($r=0.894, p<0.001$) and the ND group ($r=0.816, p<0.001$). This strong association indicates that avatar customization accounted for a substantial portion of the session, highlighting its cognitive and emotional salience within the self-as-context task.

3.2.5 Value clarification and commitment with obstacles (H5)

In Session 5, we assessed whether participants could successfully clarify their values and commit to them while encountering obstacles, reflecting the two core ACT components of values and committed action. To reflect on these components, we analyzed two key tasks: identifying personal values and navigating through challenges to pursue life goals. As shown in Fig. 3e, the time taken to identify values was significantly longer for the StD group ($Mean=18.50$ s, $SD=4.50$) than for the ND group ($Mean=16.20$ s, $SD=4.00$), showing a statistically significant difference ($F(1,50)=8.23, p=0.006, \eta_p^2=0.13$). In the navigation task, the StD group also took longer to pursue their goals while overcoming obstacles ($Mean=6.60$ min, $SD=1.50$) compared to the ND group ($Mean=6.18$ min, $SD=1.00$), with the difference being statistically significant ($F(1,50)=10.36, p=0.002, \eta_p^2=0.16$), as shown in Fig. 3f. These results suggest that the StD group experienced greater difficulty in clarifying their values and engaging in committed actions when faced with challenges, compared to the ND group.

3.2.6 Summary of hypothesis testing results

We had five hypotheses based on the key user tasks defined in accordance with the storyline of each session (Table 1). Each session was designed to evaluate different behavioral patterns between the StD and ND groups by testing the hypotheses corresponding to the unique user tasks.

In summary, all five hypotheses were supported by log and sensor data that captured distinct behavioral patterns in the StD group across different ACT components. Specifically, the StD group exhibited a heightened sensitivity to negative emotions (H1), took longer to disengage from intrusive thoughts (H2), experienced delays in self-representation

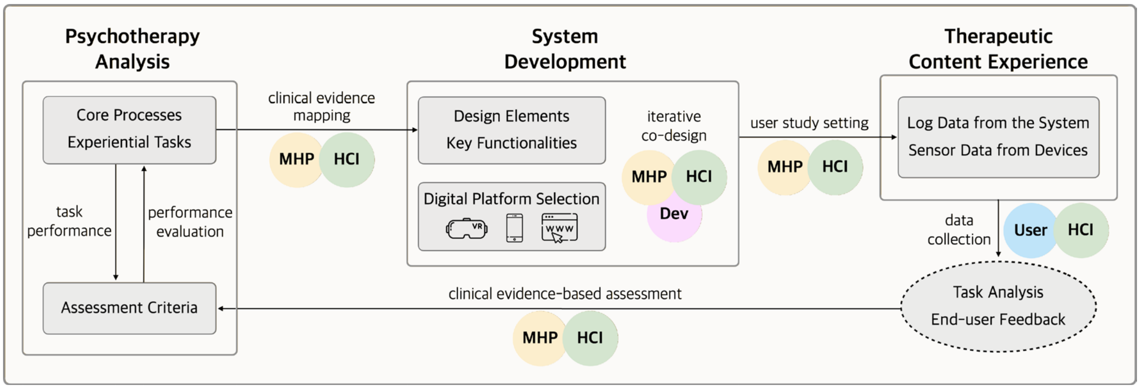

Fig. 4 The key components of the proposed framework for digital transformation in the mental health domain: (left) psychotherapy analysis, (middle) system development, and (right) therapeutic content experience. This framework includes the analysis of the selected psychotherapy, the system development process in which these elements are externalized into a digital environment, the content experience provided through user studies, and the clinical evidence-based evaluation of the collected content data. The overlapping circles represent the "collaborative" roles of HCI researchers (HCI), mental health professionals (MHPs), developers (Dev) and end-users (User), showing how their contributions intersect throughout the entire framework

tasks (H4), and required more time to clarify their values and engage in committed actions when faced with challenges (H5), compared to the ND group. Session 3 (H3) showed comparable levels of present-moment retention between the groups. Given the private and non-interactive nature of Session 3, which was specifically designed to facilitate present-moment focus, it was expected that participants with StD would exhibit attentional engagement comparable to that of the ND group. Unlike other sessions that required interaction with virtual agents or characters, Session 3 provided a private environment focused solely on present-moment engagement. Both groups maintained similarly high levels of attentional focus, suggesting that immersive and distraction-free settings (e.g., environments without virtual agents or interactive characters) may help individuals with StD perform present-focused tasks as effectively as their non-depressed peers.

These findings suggest that *VR-ACT* can be an effective tool for delivering the core principles of ACT to participants with StD and for achieving outcomes consistent with therapeutic goals. By externalizing psychotherapeutic components as interaction data, this approach also highlights the potential of digital environments to enable objective, data-driven evaluation of therapeutic engagement. This illustrates the broader promise of digital transformation in psychotherapy. In the next section, we will discuss key directions for advancing the digital transformation of psychotherapy, addressing both technical and collaborative approaches to optimizing therapeutic engagement and system personalization.

4 Discussion

4.1 Conceptual framework proposal

Based on our study results and insights, we propose a conceptual framework for digital transformation in the mental health domain, focusing on the integration of psychotherapy components into digital platforms. The framework addresses two key research areas: (1) the importance of clinical mapping to translate psychotherapy processes into measurable user interaction data to ensure that digital interventions are aligned with therapeutic goals, and (2) the role of collaboration between mental health stakeholders to develop and achieve evidence-based digital interventions. An overview of the framework is illustrated in Fig. 4, with detailed descriptions provided in Sections 4.1.1 and 4.1.2.

4.1.1 Clinical evidence mapping: Transforming psychotherapeutic components into digital interaction metrics

As mentioned in Section 3.2, our hypothesis testing based on clinical evidence confirmed that the StD group exhibited different interaction patterns during key user tasks compared to the ND group. This finding is particularly significant as it is based on the core principles presented in the ACT Hexaflex model and its associated key user tasks (Table 1). To extract potential clinical indicators from content data collected in digital environments, it is important to externalize the core processes of established psychotherapy into system design elements and key functionalities. For example, in ACT, evaluating a patient's level of acceptance or cognitive flexibility during treatment requires that these psychological characteristics be implemented as measurable interactions in the digital environment. However, simply implementing content without explicit clinical mapping and analyzing data collected from such content may not adequately serve as meaningful metrics of user interaction.

Ensuring that core psychotherapy processes are implemented as “measurable” user interaction data is a critical challenge in the digital transformation of psychotherapy. While clinical evidence in traditional therapeutic settings is primarily assessed through direct observation and patient feedback, the ability to map clinical evidence to user interaction data is essential for validating therapeutic effects in digital environments. This process, which is applicable not only to ACT but also to a wide range of psychotherapy methods, allows core processes in CBT or BAT to be externalized as digital interaction metrics that provide clinicians with quantifiable indicators of psychological progress. Thus, our framework provides a scalable model for integrating diverse psychotherapy methods into digital systems while ensuring that therapeutic intent is maintained through measurable data (see Fig. 4).

This framework enables the use of performance patterns and responses to experiential tasks as meaningful clinical indicators, allowing the evaluation of whether the core processes of the selected psychotherapy method have been appropriately implemented through the systematic collection and analysis of interaction data. By systematically mapping clinical evaluation criteria to user interactions in digital environments, user interaction data derived from digitally transformed systems can serve as objective evaluation metrics, reducing reliance on self-reports that can be subjective. This approach is essential for enabling data-driven evaluation of psychotherapy in digital environments, and can significantly support clinicians in monitoring patient progress in real-time and making informed clinical decisions. Extending this framework to other psychotherapy methods can provide a conceptual foundation for the digital transformation of mental health care, enabling consistent, data-driven evaluations across various therapeutic approaches.

4.1.2 The importance of interdisciplinary collaboration in psychotherapy digital transformation

Section 3.1 confirmed that the digital transformation of ACT effectively reflects the core principles of traditional ACT, delivered without the potential constraints associated with VR experiences. Specifically, the analysis of session-specific questions (see Table 3) developed in collaboration with MHPs, based on traditional ACT evaluation criteria and of standardized measures (IPQ, NASA-TLX, SUS) commonly used to evaluate the usability of VR systems, allowed us to validate that users experienced the digitally transformed system in alignment with our intended design framework prior to hypothesis testing. The IPQ results (See Table 4) indicated that *VR-ACT* elicited a high level of presence, defined as the psychological state in which users feel physically present in the virtual environment. Prior research has suggested that presence can positively influence restorativeness, such as stress reduction and attention restoration (Clemente et al. 2025; Dillon and Cai 2022). Building on this rationale, VR-based digitally transformed ACT may also facilitate psychological restoration and promote user engagement.

These validation results imply that systematic collaboration between HCI researchers and MHPs plays a crucial role in the digital transformation process of psychotherapy (see Fig. 4). Our framework illustrates that this process goes beyond simple technical implementation and requires the harmonious integration of clinical expertise and user experience. HCI researchers reinterpret MHPs’ clinical expertise in the context of digital platforms, transform it into measurable interaction metrics, and facilitate effective communication among stakeholders with diverse expertise. This demonstrates the potential for core elements of traditional psychotherapy to be effectively implemented in digital environments.

In this collaborative framework, the clinical expertise and continuous feedback of MHPs are essential to ensure the therapeutic validity of the system. Meanwhile, HCI researchers reconstruct the therapeutic content that incorporates both user experience and clinical evidence, interpret the collected data in a clinical context, and coordinate the collaboration among various stakeholders. Future research should explore how this collaborative approach can be applied to other psychotherapies and mental health issues, thereby extending the versatility and applicability of our framework.

4.2 Individual differences in present-moment retention and the potential of feedback loops

In Session 3, both the StD and ND groups demonstrated relatively high levels of present-moment retention during the pattern drawing and observation tasks, with no statistically significant difference between the two groups (Section 3.2.3). This finding may be attributed to the private setting of Session 3, where minimal external distractions (e.g., absence of interactive virtual agents or characters) likely facilitated user focus. The immersive nature of VR may also have provided an environment conducive to focus, allowing users who might typically struggle with present-moment tasks, such as those in the StD group, to engage more effectively than in non-immersive settings (Akinola and Mendes 2008; Bernstein et al. 2015). Nevertheless, some participants in the StD group still exhibited relatively lower levels of present-moment retention compared to the ND group (H3), highlighting the need to consider individual differences in response to immersive tasks.

The slightly lower level of present-moment retention observed in some participants with StD indicates that while the successful performance in Sessions 1–2 helped to improve present-moment retention in many participants with StD, some still struggled to maintain it during Session 3 (despite the private, non-interactive nature of the session, which did not involve engagement with virtual agents or characters). Figure 3c demonstrates that while there was no statistically significant difference in the mean number of saccades between groups, some participants exhibited difficulties maintaining focus in the present moment. Specifically, among participants with saccades above the third quartile, nine participants struggled (five of them were from the StD group) with the mean number of saccades for the StD group being 7.6 compared to 3.25 for the ND group. Additionally, two outliers in the StD group showed exceptionally high saccade values (11 and 15), indicating significant challenges with present-moment tasks. This highlights the importance of addressing individual differences, particularly in how users with depressive symptoms engage with immersive tasks, and the need for further investigation and management of those who may still struggle with present-moment retention.

One possible way to more effectively support these users with low present-moment retention is to integrate personalized feedback mechanisms between sessions. For example, providing tailored feedback based on performance in Sessions 1–2 may help users recall their previous successes in emotional regulation and cognitive defusion, thereby increasing their confidence to focus on present-moment tasks in Session 3. This approach could be particularly useful for participants with StD who continue to experience challenges with present-moment retention.

Moreover, during the session, the system can be designed to detect moments when a user's present-moment retention decreases by utilizing eye and head movements and provide adaptive feedback to maintain engagement. These feedback loops can address individual differences in present-moment retention and provide more dynamic and personalized support for users who may struggle to maintain focus during immersive therapeutic tasks.

4.3 Expanding avatar customization options to enhance the self-as-context experience

In Session 4, the time spent on avatar customization was significantly longer for the StD group compared to the ND group (Section 3.2.4). In addition, Pearson correlation analysis showed higher correlations between total session time and avatar customization time for the StD group (89.4% for the StD group and 81.6% for the ND group), suggesting that the StD group may have invested more psychological resources in this process.

Furthermore, participants were asked a session-specific question based on clinical guidelines: "Did this session help you observe yourself objectively?" (Table 3). The session-specific question scores for Session 4 showed that the StD group reported significantly lower scores than the ND group (Table 4), indicating that the StD group experienced relative difficulty with the self-as-context experience.

Although the same customization options were provided to all participants, individuals in the StD group may have found it psychologically challenging to adequately express themselves using the limited set of options (hair, face, top, and bottom). As a result, these limitations may have hindered deeper self-reflection and their engagement with the self-as-context experience. This highlights the need for more diverse and expressive customization elements.

To better encourage deeper self-reflection through the self-as-context task, it may be necessary to expand the avatar customization options. For example, offering more detailed options such as facial expressions, emotional states, posture, and physical features would allow users to engage more deeply with the task, invest more psychological resources, and thereby enhance the self-as-context experience.

Therefore, expanding avatar customization options could provide a more effective design for facilitating self-reflection and deeper task engagement. This would allow users to have a more meaningful self-awareness experience and more clearly reflect task duration and psychological changes.

4.4 Adaptive supports to facilitate cognitive defusion and self-as-context difficulties

Sessions 2 (*cognitive defusion*) and 4 (*self-as-context*) showed that the StD group required more time, reported higher levels of workload and frustration, and achieved lower session-specific question scores than the ND group (Section 3.1). These findings indicate that distancing from intrusive thoughts and adopting an observer perspective imposed greater cognitive and emotional demands on StD participants.

A plausible underlying mechanism involves the way VR-ACT transforms internal experiences into manipulable external representations through embodied interactions (e.g., removing "music bugs" or constructing an avatar). For individuals with StD, who often exhibit elevated negative affect and cognitive rigidity (Hayes et al. 2006; Juul et al. 2019), this transformation process may introduce additional cognitive load, potentially prolonging engagement with defusion and self-representation tasks. For example, in Session 2, the act of removing music bugs required participants to access, regulate, and externalize emotionally salient content, likely contributing to longer task completion times and increased affective burden. In Session 4, constraints in avatar customization options may have limited participants' self-expression, thereby increasing demands on self-referential processing. In contrast, the absence of group differences in Session 3 (*being present*) suggests that the low-distraction, private space may have effectively supported present-moment focus for both groups. Overall, these findings highlight that while VR-ACT enables immersive and experiential engagement with ACT principles, it also surfaces StD-specific challenges during cognitively demanding tasks.

To mitigate these challenges, future iterations of VR-ACT could incorporate adaptive scaffolding strategies, such as graded defusion exercises, just-in-time prompts, expanded avatar customization, and preparatory micro-tasks preceding more abstract or emotionally intense exercises. For example, if prolonged hesitation is detected during the music bugs removal task, the system could trigger a brief guided defusion prompt to assist users in externalizing intrusive thoughts. Such features are expected to lower cognitive load, accelerate skill acquisition, and enhance engagement for users with StD.

4.5 Limitations and future work

We acknowledge the limitations of our study, which should be addressed in future research.

The first limitation is the lack of longitudinal evaluation of *VR-ACT*. Although *VR-ACT* was designed in collaboration with MHPs and each session was aligned with core therapeutic principles across Sessions 1–5, we did not assess the long-term effects of repeated exposure to *VR-ACT* on user behavior. Thus, changes in user behavior patterns over multiple cycles of *VR-ACT* remain unexplored, limiting our ability to evaluate sustained clinical efficacy and long-term usability. Additionally, the lack of longitudinal tracking and repeated outcome assessments (e.g., PHQ-9, Generalized Anxiety Disorder-7 (GAD-7) (Spitzer et al. 2006)) limits our understanding of sustained engagement and therapeutic retention. To address this, future research will aim to recruit a larger sample and implement a confirmatory clinical trial in which participants complete the full set of *VR-ACT* sessions multiple times. This design will help increase the generalizability of the behavioral observations in the StD group and strengthen evidence for the clinical relevance of *VR-ACT*. Future studies should also examine the relationship between behavioral indicators (e.g., task completion time, saccades) and subjective self-reports (e.g., IPQ, NASA-TLX) to triangulate findings and enhance the validity of psychological inferences. This repeated-session design will also allow for exploratory analyses of intra-subject trends across sessions, potentially revealing personalized trajectories of therapeutic engagement and responsiveness. In addition, the mappings between behavioral patterns and underlying psychological constructs in our study remain inherently inferential. To enhance the validity and clinical relevance of these interpretations, future research could incorporate therapist feedback or conduct qualitative user interviews as complementary validation settings. To further expand this work, we aim to systematically predict changes in depressive symptoms by collecting logs and sensor data across sessions and training a Large Language Model (LLM)-based time-series prediction model. This model will integrate interaction logs, sensor data, and hypothesis-driven prompts to deliver personalized support for individuals with StD, aiding clinicians in making informed treatment decisions.

The second limitation pertains to the age range of participants. To mitigate the potential influence of VR technology as an extraneous variable, we limited the age range of participants to young adults (18–40 years) and included those with a minimum level of digital literacy based on NDLA scores (Section 2.3). While effective in the current study, this may limit generalizability to broader age groups. Future research will gradually expand the age range to investigate the applicability of *VR-ACT* across diverse demographic groups. To address technology-related barriers in older populations, we plan to implement a structured digital literacy program that includes VR device training and basic digital skills assessment prior to *VR-ACT* sessions. This systematic onboarding process will allow us to enable the inclusion of a broader age range while preserving the therapeutic potential of *VR-ACT* experience. Recent studies have demonstrated the feasibility of VR use in older adults (Hassandra et al. 2021; Kwan et al. 2021), indicating that adapting *VR-ACT* for aging populations through tailored support and training represents a promising direction for future research.

The third limitation concerns the absence of attention validation mechanisms. Although all participants completed the *VR-ACT* sessions and successfully performed the tasks, the study did not include attention checks within psychological assessments (e.g., catch items or response consistency metrics). While the IPQ results collected after each session indicated that most participants reported high levels of immersion and focus, these scores alone may not fully ensure the reliability of individual responses. Although engagement data such as task completion logs were collected via user-system interaction data, the sequential structure of the *VR-ACT* tasks limits their effectiveness in detecting inattentiveness or disengagement. To address this issue, future studies will incorporate an attention validation procedure into both psychological assessments and interaction data to enhance the robustness and interpretability of findings.

5 Conclusion

In this paper, we presented *VR-ACT*, which provides immersive therapy for individuals with StD. Through collaboration between MHPs and HCI researchers, *VR-ACT* was designed to adapt the core principles of ACT in the design of the VR content and to leverage quantitative indicators, such as VR sensor signals and system interaction logs. The goal was to investigate whether clinical evidence could be reflected in the user interaction data. The results of a quasi-experimental study with 56 participants (28 StD group, 28 ND group) demonstrated that all five hypotheses, developed based on clinical evidence and key user tasks in each session, were supported. These results provide empirical support for the feasibility of *VR-ACT* as a digital system for mental health support, and highlight distinct behavioral patterns in the StD group that can serve as clinical indicators. These findings highlight the role of immersive digital environments in supporting therapeutic interventions such as ACT. Furthermore, our proposed framework emphasizes mapping clinical evidence to user interaction data and fostering interdisciplinary collaboration led by HCI researchers. This framework provides a strategic direction for achieving the digital transformation of mental health systems, with potential application to various psychotherapies and mental health contexts.

Author contributions All authors contributed to the conceptualization and project administration of the study. B.K., D.J., Y.C., and K.H. were involved in data curation, formal analysis, investigation, methodology development, validation, visualization, and writing the original draft. H.K. and K.H. acquired the funding. B.K., D.J., and K.H. participated in reviewing and editing the manuscript.

Funding This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2020-II201373, Artificial Intelligence Graduate School Program(Hanyang University)).

Data availability The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare they have no Conflict of interest and Conflict of interest.

Ethics approval and consent to participate The study was reviewed and approved by the institutional review board at the authors' institution (No. HYU-2023-122).

Consent for publication Informed consent for publication was obtained from all individuals included in the study.

References

Akinola M, Mendes WB (2008) The dark side of creativity: biological vulnerability and negative emotions lead to greater artistic creativity. Pers Soc Psychol Bull 34(12):1677–1686

Bach PA, Moran DJ (2008) ACT in practice: case conceptualization in acceptance and commitment therapy. New Harbinger Publications

Bernstein A, Hadash Y, Lichtash Y, Tanay G, Shepherd K, Fresco DM (2015) Decentering and related constructs: a critical review and metacognitive processes model. Perspect Psychol Sci 10(5):599–617

Bertha EA, Balázs J (2013) Subthreshold depression in adolescence: a systematic review. Eur Child Adolesc Psychiatry 22:589–603

Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, Asmundson GJ, Carlbring P, Powers MB (2019) Virtual reality exposure therapy for anxiety and related disorders: a meta-analysis of randomized controlled trials. J Anxiety Disord 61:27–36

Cho JW, Kim H-W, Jeong J, Baek K (2023) Insomnia treatment in the digital healthcare era: present and future. J Sleep Med 20(2):81–90

Cieślik B, Juszko K, Kiper P, Szczepańska-Gieracha J (2023) Immersive virtual reality as support for the mental health of elderly women: a randomized controlled trial. Virtual Reality 27(3):2227–2235

Clemente D, Theodorou A, Romano L, Russo C, Rodelli R, Casagrande G, Panno A (2025) The effect of exposure to vr vs. 2d virtual environments on restorativeness: the mediating role of the sense of presence. Int J Hum Comput Interact 41(8):4567–4582

Cuijpers P, Smit F (2004) Subthreshold depression as a risk indicator for major depressive disorder: a systematic review of prospective studies. Acta Psychiatr Scand 109(5):325–331

Dillon D, Cai J (2022) Virtual reality greenspaces: does level of immersion affect directed attention restoration in VR environments? J 5(3):334–357

Dwivedi R, Mehrotra D, Chandra S (2022) Potential of internet of medical things (IoMT) applications in building a smart healthcare system: a systematic review. J Oral Biol Craniofac Res 12(2):302–318

Farah WH, Alsawas M, Mainou M, Alahdab F, Farah MH, Ahmed AT, Mohamed EA, Almasri J, Gionfriddo MR, Castaneda-Guarderas A (2016) Non-pharmacological treatment of depression: a systematic review and evidence map. Evid Based Med 21(6):214–221

Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, Nguyen QV (2019) What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 10:759

Gartlehner G, Wagner G, Matyas N, Titscher V, Greimel J, Lux L, Gaynes BN, Viswanathan M, Patel S, Lohr KN (2017) Pharmacological and non-pharmacological treatments for major depressive disorder: review of systematic reviews. BMJ Open 7(6):014912

Gibb BE, McGeary JE, Beevers CG (2016) Attentional biases to emotional stimuli: key components of the RDoC constructs of sustained threat and loss. Am J Med Genet B Neuropsychiatr Genet 171(1):65–80

Gu Q, Zhang Q, Han Y, Li P, Gao Z, Shen M (2024) Microsaccades reflect attention shifts: a mini review of 20 years of microsaccade research. Front Psychol 15:1364939

Hao X, Jia Y, Chen J, Zou C, Jiang C (2023) Subthreshold depression: a systematic review and network meta-analysis of non-pharmacological interventions. Neuropsychiatr Dis Treat. https://doi.org/10.2147/NDT.S425509

Harris, R. (2022). The happiness trap: How to stop struggling and start living. Boston, MA: Shambhala Publications.

Hassandra M, Galanis E, Hatzigeorgiadis A, Goudas M, Mouzakidis C, Karathanasi EM, Petridou N, Tsolaki M, Zikas P, Evangelou G (2021) A virtual reality app for physical and cognitive training of older people with mild cognitive impairment: mixed methods feasibility study. JMIR Serious Games 9(1):24170

Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J (2006) Acceptance and commitment therapy: model, processes and outcomes. Behav Res Ther 44(1):1–25

Henderson C, Evans-Lacko S, Thornicroft G (2013) Mental illness stigma, help seeking, and public health programs. Am J Public Health 103(5):777–780

Inc OAP (2024) Rejoyn: Computerized behavioral therapy device for psychiatric disorders. U.S, Food and Drug Administration

Jeuring H, Huisman M, Comijs H, Stek M, Beekman A (2016) The long-term outcome of subthreshold depression in later life. Psychol Med 46(13):2855–2865

Jiang X, Luo Y, Chen Y, Yan J, Xia Y, Yao L, Wang X, He S, Wang F, Wang T (2021) Comparative efficacy of multiple therapies for the treatment of patients with subthreshold depression: a systematic review and network meta-analysis. Front Behav Neurosci 15:755547

Jin S, Kim B, Han K (2025) I Don't Know Why I Should Use This App": Holistic Analysis on User Engagement Challenges in Mobile Mental Health. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems pp. 1–23

Judd LL, Rapaport MH, Paulus MP, Brown JL (1994) What is subsyndromal symptomatic depression (SSD)? Past, present and future of psychiatry: (in 2 volumes). World Scientific, Singapore, pp 295–299

Juul S, Poulsen S, Lunn S, Sørensen P, Jakobsen JC, Simonsen S (2019) Short-term versus long-term psychotherapy for adult psychiatric disorders: a protocol for a systematic review with meta-analysis and trial sequential analysis. Syst Rev 8:1–13

Kammler-Sücker KI, Löffler A, Flor H (2023) Effects of personalized movement models in virtual reality on pain expectancy and motor behavior in patients with chronic back pain: a feasibility study. Virtual Reality 27(4):3581–3598

Kim B, Jeong D, Kim JG, Hong H, Han K (2023) V-DAT (virtual reality data analysis tool): Supporting self-awareness for autistic people from multimodal VR sensor data. In: Proceedings of the 36th annual ACM symposium on user interface software and technology, pp 1–13

Kim B, Jeong D, Choi Y, Choi Y, Kim H, Han K (2025) Enhancing mindfulness-based cognitive therapy in a virtual reality: A prospective interventional study. Scientific Reports 15(1):29534.

Kim B, Jeong D, Hong H, Han K (2024) Narrating routines through game dynamics: Impact of a gamified routine management app for autistic individuals. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, pp. 1–15

Kim B, Jeong D, Jeong M, Noh T, Kim S-I, Kim T, Jang S-Y, Yoo HJ, Kim J, Hong H (2022) VISTA: User-centered VR training system for effectively deriving characteristics of people with autism spectrum disorder. In: Proceedings of the 28th ACM symposium on virtual reality software and technology, pp 1–12

Kim B, Lee D, Min A, Paik S, Frey G, Bellini S, Shih PC (2020) PuzzleWalk: A theory-driven iterative design inquiry of a mobile game for promoting physical activity in adults with autism spectrum disorder. Plos one, 15(9):e0237966

Kim SI, Jang SY, Kim T, Kim B, Jeong D, Noh T, Kim JG (2024) Promoting self-efficacy of individuals with autism in practicing social skills in the workplace using virtual reality and physiological sensors: Mixed methods study. JMIR Formative Research, 8:e52157

Kober SE, Neuper C (2013) Personality and presence in virtual reality: does their relationship depend on the used presence measure? Int J Hum Comput Interact 29(1):13–25

Kwan RYC, Liu JYW, Fong KNK, Qin J, Leung PK-Y, Sin OSK, Hon PY, Suen LW, Tse M-K, Lai CK (2021) Feasibility and effects of virtual reality motor-cognitive training in community-dwelling older people with cognitive frailty: pilot randomized controlled trial. JMIR Serious Games 9(3):28400

Lee D, Frey GC, Min A, Kim B, Cothran DJ, Be S, Shih PC (2020) Usability inquiry of a gamified behavior change app for increasing physical activity and reducing sedentary behavior in adults with and without autism spectrum disorder. Health informatics journal 26(4):2992-3008

Li J, Theng Y-L, Foo S (2014) Game-based digital interventions for depression therapy: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 17(8):519–527

Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M (2019) The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 18(3):325–336

Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, Karr CJ, Kaiser SM, Rossom RC, Bardsley LR (2017) Intellicare: an eclectic, skills-based app suite for the treatment of depression and anxiety. J Med Internet Res 19(1):10

Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, Pulkki-Råback L, Baumeister H, Sander LB (2021) Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 147(8):749

Noyes B, Biorac A, Vazquez G, Khalid-Khan S, Munoz D, Booij L (2023) Eye-tracking in adult depression: protocol for a systematic review and meta-analysis. BMJ Open 13(6):069256

Phelan I, Furness PJ, Matsangidou M, Carrion-Plaza A, Dunn H, Dimitri P, Lindley SA (2023) Playing your pain away: designing a virtual reality physical therapy for children with upper limb motor impairment. Virtual Reality 27(1):173–185

Rayner K (1998) Eye movements in reading and information processing: 20 years of research. Psychol Bull 124(3):372

Reese G, Stahlberg J, Menzel C (2022) Digital Shinrin–yoku: do nature experiences in virtual reality reduce stress and increase well-being as strongly as similar experiences in a physical forest? Virtual Reality 26(3):1245–1255

Riedl MO, Young RM (2006) From linear story generation to branching story graphs. IEEE Comput Graph Appl 26(3):23–31

Rodríguez MR, Nuevo R, Chatterji S, Ayuso-Mateos JL (2012) Definitions and factors associated with subthreshold depressive conditions: a systematic review. BMC Psychiatry 12:1–7

Solomon A, Haaga DA, Arnow BA (2001) Is clinical depression distinct from subthreshold depressive symptoms? A review of the continuity issue in depression research. J Nerv Ment Dis 189(8):498–506

Spitzer RL, Kroenke K, Williams JB, Löwe B (2006) A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 166(10):1092–1097

Spyridonis F, Daylamani-Zad D, Nightingale J (2024) Publicvr: a virtual reality exposure therapy intervention for adults with speech anxiety. Virtual Real 28(2):105

Sun Y, Ji M, Leng M, Li X, Zhang X, Wang Z (2022) Comparative efficacy of 11 non-pharmacological interventions on depression, anxiety, quality of life, and caregiver burden for informal caregivers of people with dementia: a systematic review and network meta-analysis. Int J Nurs Stud 129:104204

Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, Carvalho AF, Keshavan M, Linardon J, Firth J (2021) The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 20(3):318–335

Wang D, Shi L, Li L, Guo X, Li Y, Xu Y, Yin S, Wu Q, Yang Y, Zhuang X (2018) Subthreshold depression among diabetes patients in Beijing: cross-sectional associations among sociodemographic, clinical, and behavior factors. J Affect Disord 237:80–86

Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, Chen-Li D, Iacobucci M, Ho R, Majeed A (2020) Impact of covid-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 277:55–64

Xu W, Liang H-N, Baghaei N, Ma X, Yu K, Meng X, Wen S (2021) Effects of an immersive virtual reality exergame on university students' anxiety, depression, and perceived stress: pilot feasibility and usability study. JMIR Serious Games 9(4):29330

Zhang R, Peng X, Song X, Long J, Wang C, Zhang C, Huang R, Lee TMC (2023) The prevalence and risk of developing major depression among individuals with subthreshold depression in the general population. Psychol Med 53(8):3611–3620

Zhu K (2022) Immersive virtual reality-based cognitive intervention for the improvement of cognitive function, depression, and perceived stress in older adults with mild cognitive impairment and mild dementia: pilot pre-post study. JMIR Serious Games 10(1):32117