

User preferences for game design elements in digital mental health interventions

Shanshan Hu , Manuel Schmidt-Kraepelin , Simon Krohmann , Scott Thiebes & Ali Sunyaev

To cite this article: Shanshan Hu , Manuel Schmidt-Kraepelin , Simon Krohmann , Scott Thiebes & Ali Sunyaev (27 Sep 2026): User preferences for game design elements in digital mental health interventions, Behaviour & Information Technology, DOI: [10.1080/0144929X.2026.2736084](https://doi.org/10.1080/0144929X.2026.2736084)

To link to this article: <https://doi.org/10.1080/0144929X.2026.2736084>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 27 Sep 2026.



Submit your article to this journal [↗](#)



Article views: 157



View related articles [↗](#)



View Crossmark data [↗](#)



User preferences for game design elements in digital mental health interventions

Shanshan Hu^a, Manuel Schmidt-Kraepelin^a, Simon Krohmann^b, Scott Thiebes^c and Ali Sunyaev^a

^aSchool of Computation, Information and Technology, Technical University of Munich, Heilbronn, Germany; ^bInstitute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Karlsruhe, Germany; ^cSchool of Economics and Management, Advanced Institute of Business, Tongji University, Shanghai, People's Republic of China

ABSTRACT

Gamification is a promising strategy to enhance user engagement in digital mental health interventions (DMHIs), but its impact depends strongly on the employed game design elements (GDEs), as well as their alignment with users and contexts. The first step for guiding the selection of GDEs for DMHIs is a deep understanding of user preferences in this sensitive domain, which remains underexplored. In line with extant research investigating users' GDE preferences in other domains, we conducted scenario-based online surveys using the Best-Worst-Scaling approach with 648 potential users across five DMHI application contexts. Results reveal consistent preferences for performance-oriented elements like Goals and Progress. Compared to other domains, Stories emerged as an element with notably higher acceptance in DMHIs. Our thematic analysis of participants' justifications for their GDE preferences identified seven key rationales that explain these preferences, grouped into function-, intervention-, and user-related categories. These findings advance our understanding of gamification in DMHIs by highlighting its contextual particularities, the ambivalence of certain GDEs, and the layered nature of preference formation, thereby offering theoretical insights and practical guidance for context-sensitive and user-centered gamification design.

ARTICLE HISTORY

Received 11 September 2025
Accepted 12 September 2026

KEYWORDS

Digital mental health interventions; gamification; game design elements; best-worst-scaling; user preferences; task-technology fit

1. Introduction

Digital mental health interventions (DMHIs) have emerged as promising solutions to complement in-person therapy by offering cost-effective, accessible, and flexible mental health support (Lattie, Stiles-Shields, and Graham 2022). Results from randomised controlled trials demonstrate that DMHIs can deliver clinical improvements across various mental health conditions (Richards and Richardson 2012), including depression (Andrews et al. 2018), (social) anxiety (Guo et al. 2021), eating disorders (Linardon et al. 2020), or insomnia (Soh et al. 2020). Despite their potential, DMHIs face hurdles in real-world application, including low user engagement and adherence (Burbach et al. 2022; Gan et al. 2021). To address these challenges, researchers have begun incorporating interactive approaches into DMHIs (Gan et al. 2021), among which gamification has attracted substantial interest (Cheng et al. 2019; Johnson et al. 2016). Broadly speaking, gamification integrates elements typically found in video games (i.e. game design elements; GDEs) into non-game systems to provide their users with enjoyable

experiences while retaining the target systems' instrumental functions (Deterding et al. 2011; Liu, Santhanam, and Webster 2017). In DMHIs, gamification is typically implemented to sustain user participation, promote behaviour change, and increase the likelihood of continued use over time (Cheng et al. 2019).

Despite the growing interest in gamification for DMHIs, extant research has raised concerns that gamification may fall short of intended motivational outcomes or, in some cases, even result in unintended effects (Cheng et al. 2019). These concerns are primarily attributed to the potential misalignment between the integrated GDEs and the contextual user needs. Commonly used GDEs such as *Points*, *Badges*, and *Leaderboards* tend to emphasise achievement or social comparison, which may not align well with the needs of individuals experiencing mental distress (Cheng et al. 2019; Koivisto and Hamari 2019). Such performance- or comparison-oriented mechanisms can introduce pressure to perform and encourage comparison with others, which may be particularly counterproductive to the calm, non-judgmental, and self-focused

CONTACT Shanshan Hu shanshan.hu@tum.de School of Computation, Information and Technology, Technical University of Munich, Heilbronn, Germany

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

experience that many DMHIs aim to provide (Cheng et al. 2019). Moreover, user preferences influence how individuals perceive and respond to GDEs, thereby shaping the nature and quality of user-system interactions (Cheng et al. 2019; Koivisto and Hamari 2019). If users dislike or reject a GDE, they are less likely to engage with it as intended, reducing its potential to support sustained participation and behaviour change. As psychological states, needs for personal growth, and engagement patterns vary substantially across their users, DMHIs require a stronger alignment between GDEs and user preferences than general applications.

To advance the effective use of gamification in DMHIs, it is therefore essential to understand user preferences regarding GDEs. Several studies have explored users' GDE preferences across domains, such as healthy nutrition (Berger and Jung 2024), physical activity (Schmidt-Kraepelin et al. 2019), or learning management systems (Schöbel, Söllner, and Leimeister 2016). However, these insights may not be fully transferable to DMHIs, which are often designed for aspects like calmness or nonjudgment and require unique user-system interactions. In addition, prior studies often focus on which GDEs are favoured, but do not cover the mechanisms that give rise to these preferences (e.g. Berger and Jung 2024; Schmidt-Kraepelin et al. 2019; Schöbel, Söllner, and Leimeister 2016). Without uncovering these mechanisms, we lack insights into how these elements activate the desired motivational or self-regulatory processes in DMHIs, and cannot identify the boundary conditions under which they help or harm (Koivisto and Hamari 2019). This explanatory depth is critical, as it can explain the internal processes that connect gamification design to cognitive and behavioural outcomes, thereby clarifying why certain GDEs are effective or counterproductive in specific settings. Building on this premise, we ask the following research questions (RQs):

RQ1: Which GDEs do users prefer in DMHIs?

RQ2: What rationales explain users' GDE preferences in different DMHIs?

To answer our RQs, we conducted five online surveys informed by the best-worst scaling (BWS) approach with 648 potential DMHI users across five application contexts: mindfulness practice, sleep tracking, general mental health counselling, mood tracking, and eating disorder recovery. Our results reveal a largely consistent distribution of preferences across the five contexts. Commonly used performance-oriented elements, such as *Progress* and *Goals*, consistently received high

preference, whereas competition-oriented elements, such as *Leaderboards*, were less favoured. Despite this overall consistency, we also observed significant contextual variations rooted in the specific nature of the user-intervention interaction. By analyzing participants' justifications, we identified seven influential factors for users' GDE preferences, which we grouped into three categories (i.e. function-related, intervention-related, and user-related rationales). These insights help explain the observed distribution of rankings and advance our understanding of how users evaluate gamification in DMHIs. Drawing on these findings and grounded in task-technology fit (TTF) theory (Goodhue and Thompson 1995), we propose a multi-level fit perspective that explains the observed preference patterns and provides a foundation for research on more context-sensitive and user-centered gamified DMHIs.

The contribution of our work is threefold. First, we provide a comparative overview of users' GDE preferences across five application contexts of DMHIs. This advances the understanding of meaningful and context-sensitive gamification design in mental health, showing both common patterns (e.g. consistently high preference for *Progress* and *Goals*, low preference for *Leaderboards*) and intervention-specific nuances (e.g. low preference for *Avatars* in mindfulness practice or for *Social Interaction* in sleep tracking). In this way, our study (i) offers a practical, evidence-based baseline for designers to select effective GDEs and avoid common pitfalls, and (ii) establishes an empirical foundation that can guide future research. Second, by identifying and grouping seven rationales explaining users' GDE choices, we advance the scientific knowledge on the determinants that drive the formation of users' GDE preferences. This knowledge forms a building block for the future development of highly efficient personalised and tailored gamification concepts. Instead of simple preference matching, our findings enable the development of adaptive systems that can infer a user's underlying need in a specific application context and offer the right GDE at the right time. Lastly, we synthesise these findings into a multi-level fit perspective for gamified DMHIs. This perspective contextualises the TTF for this sensitive domain and demonstrates the layered interplay between function-, intervention-, and user-related factors. Rather than assuming a single dominant source, our model highlights that the perceived appropriateness of a GDE depends on the simultaneous alignment of its functional affordances with the goals of the intervention and users' individual needs. In doing so, our model advances our understanding of user

preference formation and provides a structured basis for future empirical studies that investigate the mechanisms by which individual GDEs operate in mental health contexts and their effectiveness.

2. Background

2.1. Gamified digital mental health interventions

Over the past 15 years, gamification has become a prominent topic within human–computer interaction and related fields such as information systems (IS), marketing, psychology, and medical informatics (Lowry, Dinev, and Willison 2017). One of the earliest and still most frequently cited definitions is provided by Deterding et al. (2011), who defined gamification as ‘the use of game design elements in non-game contexts’ (10). Subsequently, scholars extended this definition by emphasising intended outcomes like motivation (Fitz-Walter, Tjondronegoro, and Wyeth 2011), engagement (Mekler et al. 2013), or gameful experiences (Huotari and Hamari 2017). A common theme that emerges from all the relevant definitions of gamification is that gamified systems are neither solely hedonic (i.e. pleasure-oriented systems that provide self-fulfilled values to users, such as video games) nor solely utilitarian (i.e. productivity-oriented systems that provide instrumental value to users, such as decision-support systems) (Liu, Santhanam, and Webster 2017). Rather, in gamified IS, hedonic and utilitarian aspects converge and need to be carefully balanced to ensure effective use (Koivisto and Hamari 2019). In this work, we adopt the understanding and terminology by Liu, Santhanam, and Webster (2017) and follow the notion that gamified IS have a dual purpose of evoking both instrumental and experiential outcomes within their users.

In comparison to the most popular application domains of gamification, such as physical activity or education, literature on gamification in DMHIs is still relatively scarce (Cheng 2020). Scholars have attributed this to some outcomes of gamification, such as feelings of competition, social comparison, or extrinsic rewards, being inappropriate in circumstances where users may experience psychological distress (Cheng et al. 2018). Nevertheless, the adoption of gamification in DMHIs is increasing, with a recent systematic review reporting a median of five GDEs per app (Cheng et al. 2019). Moreover, empirical studies on gamification for mental health are now emerging across diverse mental-health contexts, including anxiety (Jingili et al. 2023), mood regulation (Litvin et al. 2020), eating disorders (Tang et al. 2022), substance misuse (Zhang et al. 2019),

sleep (Liang et al. 2024), general well-being (Lehtoranta, Xi, and Hamari 2024), and serious mental illnesses (Fitzgerald and Ratcliffe 2020). In DMHIs, instrumental outcomes are typically aligned with the intervention’s therapeutic objective. For instance, in a DMHI for mindfulness practice, the instrumental outcome might center on enhancing attentional control through guided meditation exercises, while a mood tracking DMHI could prioritise identifying emotional triggers by correlating logged feelings with daily activities. Experiential outcomes, by contrast, stem primarily from the implemented GDEs: goals provide orientation, whereas narratives or avatars can engender immersion and flow (Koivisto and Hamari 2019).

Although the literature broadly agrees that achievement- and competition-oriented designs may be ill-suited to mental-health contexts (Cheng et al. 2019), little is known about which specific gamification designs are most appropriate. To address similar questions, research in other application domains has shown that a critical first step is to investigate users’ GDE preferences (e.g. Berger and Jung 2024; Schmidt-Kraepelin et al. 2019; Schöbel, Söllner, and Leimeister 2016). This approach is grounded in the idea that effective gamification design requires a careful selection of GDEs as its foundation (Liu, Santhanam, and Webster 2017), which must be guided by user preferences that influence the intervention’s acceptance, feasibility, and long-term sustainability (Morschheuser et al. 2018). Accordingly, the present study takes this essential first step within the DMHI domain by investigating which GDEs potential users prefer to see implemented in DMHIs and why.

2.2. Users’ game design element preferences

User preferences are comparative evaluative judgments through which individuals express a stronger subjective inclination toward one alternative relative to others (Hertzum 2025). Such preferences are fundamental for system design and adoption because they inform how users experience and evaluate a system. User experience, in turn, shapes whether users perceive a system as acceptable and worth continued use (Hertzum 2025). In domains where sustained engagement and responsible design are essential, such as DMHIs, understanding users’ preferences regarding specific design strategies (e.g. integration of GDEs) is particularly important (Cheng et al. 2019). If GDEs are misaligned with users’ preferences or contextual expectations, they may undermine acceptance despite their intended motivational benefits (Cheng et al. 2019; Koivisto and Hamari 2019). Investigating users’ GDE preferences

thus constitutes a foundational step in identifying which gamification designs are perceived as appropriate and capable of supporting sustained adoption within the DMHI domain.

To capture users' GDE preferences, studies have employed various methods, including focus groups (Nour, Rouf, and Allman-Farinelli 2018), questionnaires (Fitz-Walter et al. 2013), collecting log data (Van Houdt et al. 2020), or user profile data (Li et al. 2019). While these design-oriented studies provide valuable first insights for developing personalised and context-specific gamification (Schöbel et al. 2021), the generalizability of their findings is limited, as they typically focused on narrowly defined systems and target audiences. Regarding more generalizable knowledge on users' GDE preferences, we identified two research streams that are related to our work.

A first stream of related literature places users and their unique characteristics at the center of inquiry. Research in this stream has attempted to cluster users into user types with similar preferences based on various factors, such as personality traits (Tondello et al. 2016), gaming habits (Hassan et al. 2020), age (Yuan and Guo 2021), or gender (Koivisto and Hamari 2014). However, despite their prominence in research and practice, gamification user typologies have received severe criticism regarding their validity, reliability, and overall usefulness for designing gamified systems (Kirchner-Krath et al. 2024). For example, a study by Hassan et al. (2020) examined the relationship between player types and user preferences regarding general gamification features (i.e. achievement, immersion, and social interaction). They found that player types are not a comprehensive predictor for gamification preferences, especially for social and immersion-related GDEs. While this study advances the understanding of the relationship between player types and gamification preferences, the authors emphasise the need for more research investigating how users form gamification preferences. Moreover, Kirchner-Krath et al. (2024) provide a comprehensive overview of user typologies that have been used in gamification research. They find that user typologies, despite their different origins, show considerable overlap, some being consistent, whereas others contradict theoretically assumed relationships. Consequently, they conclude that gamification research should shift the focus from developing and applying ever more nuanced typologies to understanding and measuring the key underlying determinants of user preferences in gamified systems.

A second stream of literature has focused on application context as a potential determinant of users' GDE preferences and has examined its influence,

independent of specific systems or user groups. Studies have examined various contexts, such as physical activity (Schmidt-Kraepelin et al. 2019), nutrition (Berger and Jung 2024), personal travel (Verzosa et al. 2018), stress management (Berger et al. 2023), and learning management systems (Schöbel, Söllner, and Leimeister 2016). Across these contexts, achievement-oriented GDEs like *Progress* and *Levels* rank among the most preferred GDEs, while immersion-oriented elements like *Avatars* and *Stories* usually rank lower. However, because each study investigates a distinct set of GDEs, synthesising findings beyond the most ubiquitous elements (e.g. *Points*, *Badges*, *Leaderboards*) remains difficult. Even when some contexts appear similar to popular DMHI application contexts (e.g. stress management and nutrition), the diversity of DMHI application contexts makes it challenging to draw clear conclusions from this body of work for the DMHI domain. Moreover, most of these studies are limited to investigating what users may prefer, without deeply exploring the underlying reasons for those preferences. Consequently, we still lack a nuanced understanding of users' GDE preferences in the DMHI domain. In our work, we aim to bridge both research streams. While methodologically aligning close to the second stream to capture the specificities of DMHIs, we simultaneously aim to investigate the context- and user-specific influential factors for users' GDE preferences by employing the approach outlined in the following.

3. Study design

To approach our RQs, we conducted scenario-based online surveys guided by the best-worst-scaling (BWS) approach (Louviere et al. 2013) across five distinct DMHI application contexts. The rigor and appropriateness of this study design are grounded in two considerations: the methodological suitability of BWS for eliciting comparative preferences among multiple design alternatives, and the established use of BWS in prior studies examining users' GDE preferences (e.g. Berger and Jung 2024; Schmidt-Kraepelin et al. 2019; Schöbel, Söllner, and Leimeister 2016).

BWS is grounded in Random Utility Theory (McFadden 1972), which posits that individuals assign a perceived utility to each option and select the one with the highest perceived utility while avoiding the one with the lowest. In the BWS approach, participants are repeatedly presented with choice sets consisting of at least three options of interest, in our case, GDEs. For each set, participants select one 'best' and one 'worst' option, which together denote the greatest difference in participants' perceived utility within the set.

Compared to direct ranking mechanisms or rating scales (e.g. Likert scales), BWS offers two key advantages (Louviere et al. 2013): (1) its forced-choice nature ensures the discrimination of options; (2) it is scale-free and avoids potential biases as participants do not need to maintain consistency in calibrating the scale across options. These advantages make BWS well-suited to investigating users' GDE preferences. In particular, the forced-choice nature reduces the likelihood of similarly high ratings for multiple GDEs, as such motivational features may all seem appealing when evaluated in isolation. Furthermore, integrating GDEs into DMHIs inherently involves design choices among alternative GDEs, as it is not feasible to implement all elements simultaneously due to limited design space and the potential to overwhelm users (Cheng et al. 2019). Consequently, the relative importance of elements identified by the BWS approach matters more than their absolute endorsement in isolation, making BWS particularly suitable for this study.

Several studies have recognised the methodological advantages of BWS and have successfully employed scenario-based online surveys informed by BWS to examine users' GDE preferences (e.g. Berger and Jung 2024; Schmidt-Kraepelin et al. 2019; Schöbel, Söllner, and Leimeister 2016). By aligning our methodological design with this stream of research, we aim to make our findings comparable and contribute to discussions on the context sensitivity of gamification. In the following, we first introduce our research contexts before delving into data collection and data analysis procedures.

3.1. Research contexts

To investigate GDE preferences in the domain of DMHIs, we selected five distinct DMHI application contexts, namely (1) mindfulness practice, (2) sleep tracking, (3) general mental health counselling, (4) mood tracking, and (5) eating disorder recovery. We chose these five contexts for two major reasons. First, they represent contexts where gamification is often implemented as part of the broader goal of improving mental health and well-being (Cheng et al. 2019). This ensures the ecological validity of our study, as participants are likely to perceive the inclusion of GDEs as plausible within these scenarios. Second, while all five contexts fit the broader label of DMHIs, they exhibit substantial variation. To make this variation explicit, we systematically characterise the contexts along several dimensions commonly used to distinguish DMHIs (Hopkin et al. 2025; Mohr et al. 2014). Specifically, we compare the contexts with respect to their primary

purpose, target behaviours, interaction frequency, target users, and topic sensitivity. Table 1 summarises the five contexts systematically among these dimensions together with a representative DMHI for each.

The *primary purpose* ranges from skill practice and self-monitoring to therapeutic support and recovery. The *target behaviours* likewise vary in complexity, from establishing a single habit (e.g. tracking sleep duration) to managing multiple interrelated behaviours (e.g. diet adherence), while *interaction frequency* ranges from once-daily logging to multiple entries per day. The contexts also address different *target users*, from non-clinical individuals seeking everyday well-being to those managing diagnosed conditions. Finally, the contexts differ in *topic sensitivity*, from highly stigmatised concerns like eating disorders to comparatively low-stigma activities like mindfulness practice. Overall, we expect this variation in contexts to translate into substantial differences in user-system interactions and users' GDE preferences.

3.2. Survey design and data collection

We decided to focus on ten GDEs as the choice options because the number of required choice sets depends on the number of options (Louviere et al. 2013), and ten items would limit the length of the survey to a reasonable length. We identified the ten GDEs by synthesising findings from literature reviews on gamification in DMHIs and from previous BWS studies on GDE preferences in health-related contexts, including nutrition (Berger and Jung 2024), stress management (Berger et al. 2023), or physical activity (Schmidt-Kraepelin et al. 2019). This set of GDEs has proven effective in prior studies, covers the most commonly implemented GDEs, and represents a diverse range of functions within gamified IS. Appendix A provides more information on the selection of GDEs. Following best practices for BWS approaches, 15 choice sets are sufficient for ten or fewer options (Orme 2009). Each participant therefore viewed 15 unique choice sets, each containing four GDEs, constructed using a balanced incomplete block design (Louviere et al. 2013; Orme 2009): each GDE appeared equally six times, and each combination of two GDEs appeared exactly twice.

Figure 1 presents our survey procedure, which was identical across contexts, with slight differences in wordings and descriptions of the GDEs noted below. Each survey started with an introduction to a target scenario, in which participants were asked to imagine using a DMHI for a specific application context (see Appendix B for details). We then presented our ten

Table 1. Overview of the five selected DMHI application contexts.

Context	Mindfulness practice	Sleep tracking	General mental health counselling	Mood tracking	Eating disorder recovery
Description	Seek to cultivate non-judgmental present-moment awareness through meditation and other practices (Sliwinski et al. 2017)	Allow to record and monitor sleep metrics and provide guidance on improving sleep quality (Baron et al. 2018)	Provide general support for the prevention and treatment of mental health issues (Firth et al. 2017)	Collect data about users' mood to identify patterns and facilitate emotional regulation (Kumar et al. 2020)	Provide information, support and treatments to individuals with eating disorders (Thomas et al. 2024)
Primary purpose	Skill practice	Self-monitoring & guidance	Therapeutic support	Self-monitoring & emotion-regulation	Treatment & recovery support
Exemplary target behaviours	Establish a consistent mindfulness meditation routine	Track sleep duration, consistency and disturbances	Complete therapeutic activities or self-reflection exercises	Log mood data regularly and reflect on emotional patterns	Build healthy eating and self-image habits
Interaction frequency	Regular practice sessions	Once per day	Session-based	Multiple entries per day	Multiple entries per day
Target users	Individuals experiencing stress	Individuals with irregular sleep schedules or sleep difficulties	Individuals undergoing counselling or therapy	Individuals with mood disorders, or interested in emotional self-knowledge	Individuals recovering from eating disorders (e.g. anorexia)
Topic sensitivity	Low	Low	Moderate-high	Low-moderate	High
Exemplary DMHI	Headspace	SleepTown	Wysa	Mood Tracker: Self-Care Habits	Recovery Record

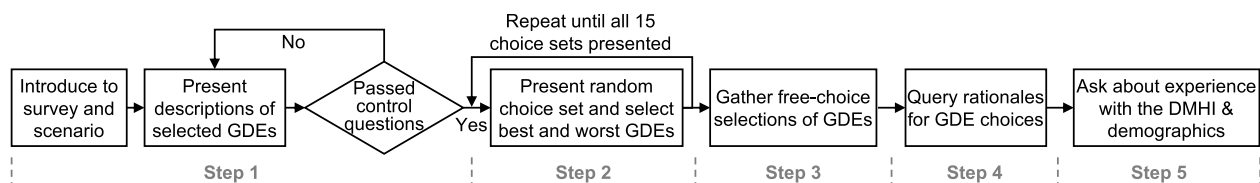
GDEs with descriptions and mockups. Across DMHIs, we aimed to keep the general style (e.g. color scheme) of the mockups consistent, while accounting for individual differences in DMHIs (e.g. target behaviours). All mockups are shown in Appendix B.

After the introduction, we included two control questions to ensure that participants understood the GDEs. Participants who passed these two questions then proceeded to complete the 15 choice sets, each time selecting their most preferred GDE (i.e. the 'best') and least preferred GDE (i.e. the 'worst'). Descriptions and mockups remained accessible throughout. Following the BWS task, we asked participants to freely choose any of the ten GDEs they wished to use (multiple selections allowed) in the current DMHI application context (i.e. free-choice selections). Afterwards, we asked participants to 'describe briefly which GDEs you like best and why'. Lastly, we collected demographic data (i.e. gender, age, and education) and asked participants to what degree they had used a DMHI in their respective context. Before rolling out each survey, we pre-tested it with experienced researchers from our network. This resulted in minor phrasing modifications and clarifications for the GDE descriptions and mockups.

3.3. Quantitative analysis

We adopted a two-stage analysis strategy for the BWS data (Louviere et al. 2013), using the R programming language and the RStudio application. First, following Flynn et al. (2007), we conducted a *counting analysis* to obtain an initial, context-specific ranking of GDEs. For each GDE, we counted the number of times it was chosen as most preferred (B) and least preferred (W), and computed the BWS scores ($B - W$). Then, we calculated the standardised mean by dividing the BWS scores by the product of the number of times each GDE appeared in a choice set (six in our case) and the number of participants. The standardised mean ranges from -1 to 1 , with higher values indicating stronger preference. We also assessed convergent validity by comparing the BWS scores and participants' free-choice selections. To do so, we formed a free-choice-based weight per GDE:

$FCW_j = \sum_i \frac{I_{ij}}{nfc_i}$, where $I_{ij} = 1$ if the participant i selected an element j in free-choice (0 otherwise), and nfc_i is the number of GDEs the participant i selected in their free-choice set. We then tested whether elements with higher BWS scores are also selected more often in free-choice selections by using linear regression.

**Figure 1.** Survey procedure.

Second, we employed *conditional logit regression* (CLR) to estimate the relative utilities of the GDEs based on participants' choices (McFadden 1972). According to the Random Utility Theory (McFadden 1972), the perceived utility u_j that an individual assigns to option j consists of a systematic component v_j and a random component ε_j (i.e. $u_j = v_j + \varepsilon_j$). The systematic component v_j is determined by observable attributes (i.e. $v_j = \sum \beta_i \cdot x_{ij}$, where β_i is the utility weight of attribute i , e.g. price, appearance; in our case, element-specific dummy variables) and x_{ij} is the level of attribute i for option j . The random component ε_j cannot be observed directly and is assumed in CLR to follow a Gumbel distribution (McFadden 1972). Under this assumption, the probability of choosing option j from the choice set C is given by: $P_j = \frac{e^{v_j}}{\sum_{k \in C} e^{v_k}}$. CLR estimates the coefficients β_i in v_j , which in our case reflect the relative desirability of each GDE compared to a reference element. Specifically, the model takes the form $v_j = \sum_{m=1}^9 \beta_m \cdot GDE_{mj}$, where GDE_{mj} is a dummy variable indicating whether option j (i.e. GDE j) corresponds to GDE m , and one element is omitted as the reference category. Accordingly, the estimated coefficients β_j can be interpreted as the relative utility of the element j compared to the reference element, whose utility is normalised to zero. We estimated separate models for each context first, and then pooled data from all contexts and estimated a combined CLR model incorporating interaction terms between DMHI and GDE. Specifically, the pooled model can be expressed as

$$v_j = \sum_{m=1}^9 \beta_m \cdot GDE_{mj} + \sum_{d,m} \gamma_{dm} \cdot (DMHI_d \cdot GDE_{mj}),$$

where $DMHI_d$ is a dummy variable indicating the intervention context d , and γ_{dm} captures how the relative utility of GDE m is moderated by context d . This allowed us to test whether the relative utility of individual elements varied significantly across DMHIs, revealing potential moderating effects of DMHIs' particularities on preference formation.

3.4. Thematic analysis of qualitative data

To investigate possible reasons for the variation in users' preferences, we applied thematic analysis to participants' responses to the open question at the end of the surveys. For the analysis, we followed the six phases outlined by Braun and Clarke (2006). In the first phase (familiarising with the data), one author familiarised themselves with the data by carefully reading all participant responses

and extracting the most relevant statements that reflected rationales for participants' GDE preferences. During the second phase (generating initial codes), initial codes were generated. For example, the statement '*help me the most just to see how far I've come*' was coded as *see achievements*, while '*I prefer elements which keep track of my progress by collecting points*' was coded as *track progress*. A second author then independently reviewed these initial codes, and discrepancies in coding were discussed and resolved within the author team. In the third phase (searching for themes), the author team iteratively used tables to identify relationships between initial codes and sorted them into different levels of themes (i.e. main overarching themes and sub-themes within them). Through this iterative process, the codes were refined and further organised into (sub-)themes (e.g. *see achievements* and *track progress* were merged into the theme *opportune feedback*). In the fourth phase (reviewing themes), the author team examined whether the candidate themes formed coherent patterns at the level of coded aspects and whether they adequately captured the meanings across the whole sample. This process led to refining the thematic structure by merging overlapping themes, splitting themes that were too broad, and reassigning aspects that did not clearly fit their initial theme. In the fifth phase (defining and naming themes), we defined the themes by identifying the 'essence' of what each theme is about and determining what aspect of the responses each theme captures and what they do not. Based on their definitions, we named the refined themes. The results of the thematic analysis are reported in section 4.3 (phase 6: producing the report). An illustrative example of this analysis process is presented in Figure 2.

4. User preferences for game design elements in DMHIs

4.1. Sample description

We employed a non-probability, self-selected sampling approach (Etikan, Musa, and Alkassim 2016), recruiting potential DMHI users through online channels in which the respective DMHI topics are actively encountered and discussed. For each of the five contexts, participants were recruited separately via topic-specific Facebook groups, professional networks on LinkedIn, and announcements in university-affiliated online platforms. For example, recruitment for the mood tracking context included Facebook groups such as Stress, Anxiety, Addiction, Depression: Mental & Emotional Health Support, whereas recruitment for the eating disorder recovery context included groups such as Eating

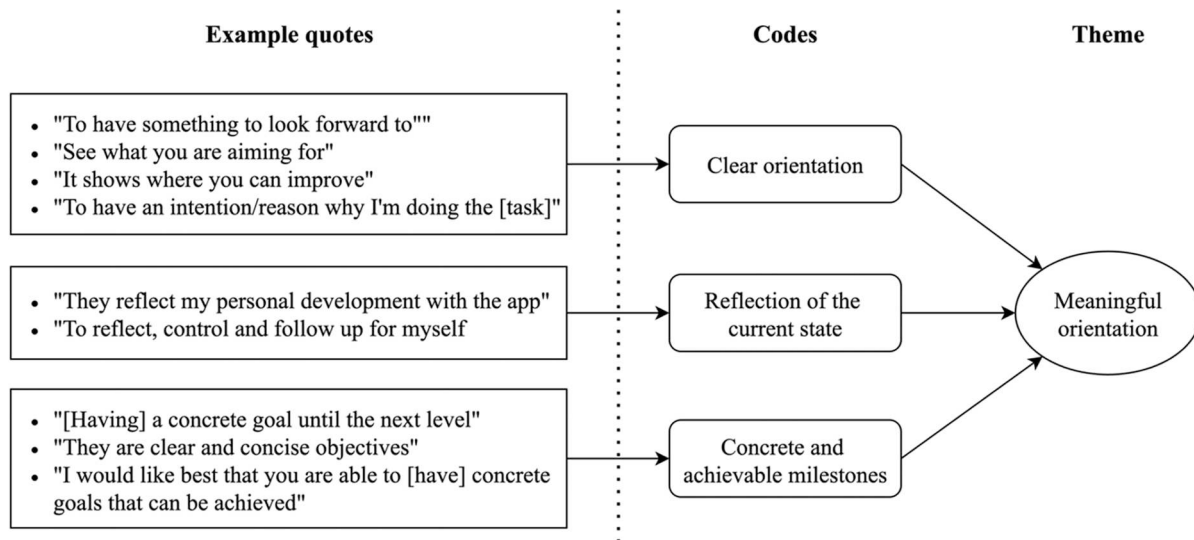


Figure 2. Illustrative examples of the thematic analysis process.

Disorder Awareness. Since our study investigates how potential users perceive and engage with DMHIs in real-world contexts, it was essential to recruit participants who are not only aware of such tools but also actively situated in the digital communities where DMHIs are encountered, evaluated, and discussed. Recruiting through these channels was therefore particularly fitting, as it ensured that participants could provide contextually grounded and experientially informed responses, yielding a sample representative of actual DMHI target users. Especially, topic-specific Facebook groups enabled access to communities already engaged with the respective DMHI topics, while LinkedIn and university-affiliated platforms broadened the participant pool to include individuals with diverse professional and academic backgrounds and varying levels of familiarity with DMHIs. Collectively, this multi-channel approach yielded a sample that is both contextually appropriate for the study's focus on DMHIs and sufficiently heterogeneous to capture a wide range of perspectives. Ethical approval for this study was not required under institutional guidelines, as it involved no intervention and collected no sensitive personal data. All participants provided informed consent before beginning the survey.

Each participant completed the survey for a single application context only. In total, 813 participants completed the questionnaire (198 for mindfulness practice, 240 for sleep tracking, 128 for general mental health counselling, 152 for mood tracking, and 95 for eating disorder recovery). Responses were excluded if participants failed the attention checks or completed the questionnaire in less than 250 s. After applying these criteria, the final datasets comprised 168 valid cases for

mindfulness practice, 181 for sleep tracking, 116 for general mental health counselling, 107 for mood tracking, and 76 for eating disorder recovery, totalling 648 valid responses. Participants ranged in age from 16 to 65 years. Gender distribution and education levels for each context are presented in Figure 3.

4.2. Preferences for game design elements across contexts

4.2.1. Counting analyses

The results of our counting analyses for all five DMHI application contexts are summarised in Table 2. Building on the method described earlier, we report a standardised mean score (S) for each GDE, which reflects the relative strength of its preference within each context. For example, in the context of mindfulness practice, 168 participants collectively selected Progress 622 times as most preferred (B) and 35 times as least preferred (W). The standardised mean was calculated as $(622 - 35) / (6 \times 168) = 0.582$. In addition, we report standard deviations (SD), which indicate the variability in preference across participants. Lower standard deviation values indicate more consistent preferences within the sample.

Across all contexts, performance-oriented elements such as *Progress* and *Goals* consistently ranked at the top, with standardised means exceeding 0.50 in multiple contexts. *Levels* and *Points* were generally rated as moderately positive, whereas *Badges* and *Stories* tended to receive near-neutral scores. In contrast, *Avatars*, *Leaderboards*, and *Virtual Goods* repeatedly appeared among the lowest-ranked elements, with negative standardised means in every context. Some variation emerged

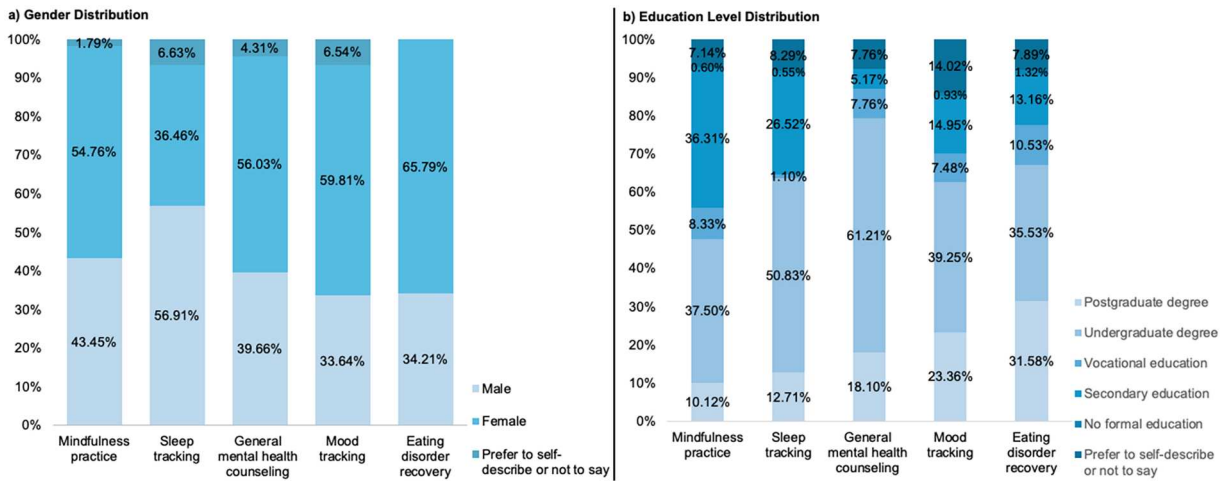


Figure 3. Gender and education level distribution across DMHI application contexts.

between contexts. For instance, *Stories* was rated close to neutral in mindfulness practice ($S = -0.055$) and mood tracking ($S = -0.047$) but slightly positive in general mental health counselling ($S = 0.148$) and sleep tracking ($S = 0.062$). Although *Social Interaction* consistently received negative evaluations, the extent varied (e.g. $S = -0.050$ in general mental health counselling vs. $S = -0.558$ in sleep tracking). Figure 4 highlights these variations by comparing the ranking positions in each context. While the broad pattern of preferences is stable, certain elements, particularly those in the bottom part of the ranking, shift notably in rank depending on the application context.

To complement these results, linear regression analyses revealed a significant positive relationship between standardised BWS scores and participants' free-choice selections across all contexts. Figure 5 visualises this relationship using scatter plots with fitted regression lines. For instance, in the mindfulness context, the regression model ($b = 26.66$, $SE = 1.64$) yielded a significant effect ($p < .001$) of the predictor variable (i.e. BWS scores), with an adjusted R^2 of .967, indicating that the model explained approximately 96.7% of the variance in participants' choices. These relationships show that the BWS-derived rankings are a very strong indicator of participants' preferences.

4.2.2. Conditional logit regression analysis

The results of the conditional logit regressions estimated separately for each application context are summarised in Table 3. These regressions capture the relative utilities of each GDE, reflecting how strongly a GDE increases or decreases the likelihood of being chosen compared to the reference GDE in the same context. In each application context, the reference is defined as the lowest-ranked GDE in the counting analysis, as we assume

that the lowest-ranked GDE has a utility of zero, with all other coefficients expressing relative deviations from this baseline. For example, in mindfulness practice, *Progress* ($\beta = 2.662$, $SE = 0.081$) and *Goals* ($\beta = 2.026$, $SE = 0.080$) were substantially more likely to be selected than the baseline element *Avatars*. A comparable pattern emerged in sleep tracking, with *Progress* ($\beta = 2.139$, $SE = 0.077$) and *Goals* ($\beta = 1.951$, $SE = 0.077$) at the top, and *Avatars* ($\beta = 0.152$, $SE = 0.097$) only marginally above the reference *Social Interaction*. The context-specific results broadly align with the counting analysis: *Progress* and *Goals* consistently occupy the top ranks, whereas *Avatars*, *Leaderboards*, and *Virtual Goods* appear at the lower end.

To examine the influence of application context on user preference, we estimated an interaction model including $GDE \times context$ terms (i.e. interaction terms). Based on the overall ranking of the ten GDEs across contexts and taking into account the relatively sensitive nature of the five application contexts, we set eating disorder recovery as the reference context and *Leaderboards* as the reference element, normalising its utility in this context to zero. Table 4 reports the estimated coefficients (Coef, i.e. relative utilities), their standard errors (SE), and significance levels ($Pr(>|z|)$) of all interaction terms.

These results capture how the utilities of individual GDEs change across contexts relative to this baseline. While some interaction effects were not statistically significant, several revealed notable contextual influences. Among the significant effects, some contexts reduced the attractiveness of particular GDEs. For instance, *Social Interaction* was positively valued in the baseline but became substantially less preferred in sleep tracking ($\beta = -1.893$, $p < .001$). Similarly, *Goals* showed a relative decline in mood

Table 2. Results of counting analyses for the five DMHI application contexts.

GDE	Mindfulness practice				Sleep tracking				General mental health counselling				Mood tracking				Eating disorder recovery			
	B	W	S	SD	B	W	S	SD	B	W	S	SD	B	W	S	SD	B	W	S	SD
Goals	463	45	0.415	0.367	549	64	0.447	0.430	315	51	0.379	0.429	274	59	0.335	0.456	272	25	0.542	0.380
Progress	622	35	0.582	0.360	602	36	0.521	0.350	283	66	0.312	0.420	363	39	0.505	0.409	266	20	0.539	0.375
Points	256	114	0.141	0.388	265	105	0.147	0.355	159	114	0.065	0.433	167	64	0.160	0.387	110	48	0.136	0.380
Levels	395	52	0.340	0.371	291	88	0.187	0.349	250	50	0.287	0.390	141	91	0.078	0.363	103	42	0.134	0.293
Badges	214	200	0.014	0.464	295	173	0.112	0.436	197	86	0.159	0.417	155	109	0.072	0.475	107	76	0.068	0.432
Stories	202	257	-0.055	0.512	313	246	0.062	0.573	218	115	0.148	0.525	136	166	-0.047	0.510	82	110	-0.061	0.474
SI	127	372	-0.243	0.508	53	659	-0.558	0.429	169	204	-0.050	0.589	142	255	-0.176	0.641	86	163	-0.169	0.585
Avatars	33	571	-0.534	0.423	82	385	-0.279	0.430	47	327	-0.402	0.466	81	250	-0.263	0.519	24	192	-0.368	0.407
VG	128	421	-0.291	0.520	155	445	-0.267	0.553	60	264	-0.293	0.412	80	265	-0.288	0.541	39	222	-0.401	0.491
LB	80	453	-0.370	0.434	110	514	-0.372	0.484	42	463	-0.605	0.453	66	307	-0.375	0.505	51	242	-0.419	0.553

SI: Social interaction; VG: Virtual goods; LB: Leaderboards.

tracking ($\beta = -0.82$, $p = .016$). At the same time, other contexts enhanced certain preferences. *Stories*, already positive at baseline, became even more valued in general mental health counselling ($\beta = 1.60$, $p < .001$), and *Avatars*, weak at baseline, turned into a positive predictor in mood tracking ($\beta = 0.95$, $p = .004$).

4.3. Rationales for users' game design element preferences in DMHIs

Across five application contexts of DMHIs, participants provided 507 valid responses to the open-ended survey question. Our thematic analysis of these responses yielded seven key rationales driving users' GDE preferences. Table 5 presents an overview of these rationales, their main mechanisms (i.e. how they influence users' GDE preferences), and exemplary GDEs. We categorised the seven rationales into three groups based on whether a rationale stems primarily from the GDE's intrinsic functionalities (*function-related*), the application context of the DMHI (*intervention-related*), or the individual's personal characteristics (*user-related*). The following sections will elaborate on these rationales in detail.

4.3.1. Function-Related Rationales

Function-related rationales are justifications for user preferences that are driven by the intrinsic function of the GDEs themselves. When participants explain their choice with these rationales, they are focusing on what the element does at a fundamental level. For instance, a preference for *Progress* reflects its inherent capacity to deliver immediate, stepwise performance information. Similarly, a preference for *Social Interaction* can be explained by the element's built-in opportunities for connection.

Meaningful Orientation. The first rationale we found was that users prefer elements that give clear direction and structure, helping them decide what to do next and why. This orientation shapes user preferences through two complementary ways: direction-setting and path-reflection. Direction-setting describes a mechanism where users choose GDEs that establish a clear forward-looking journey, which fosters a sense of purpose and motivation. This clear direction translates a broad intention (e.g. 'improve my sleep habits') into concrete, actionable steps. *Goals* are chosen because they provide 'a clear direction to work towards and make [users] very hopeful that [they] can achieve this' (ID565, sleep tracking). *Goals* allow users to 'structure [the] plan in improving' (ID1003, sleep tracking) and 'be able to be clear about what needs to be accomplished on a daily basis' (ID463, eating disorder recovery). Such

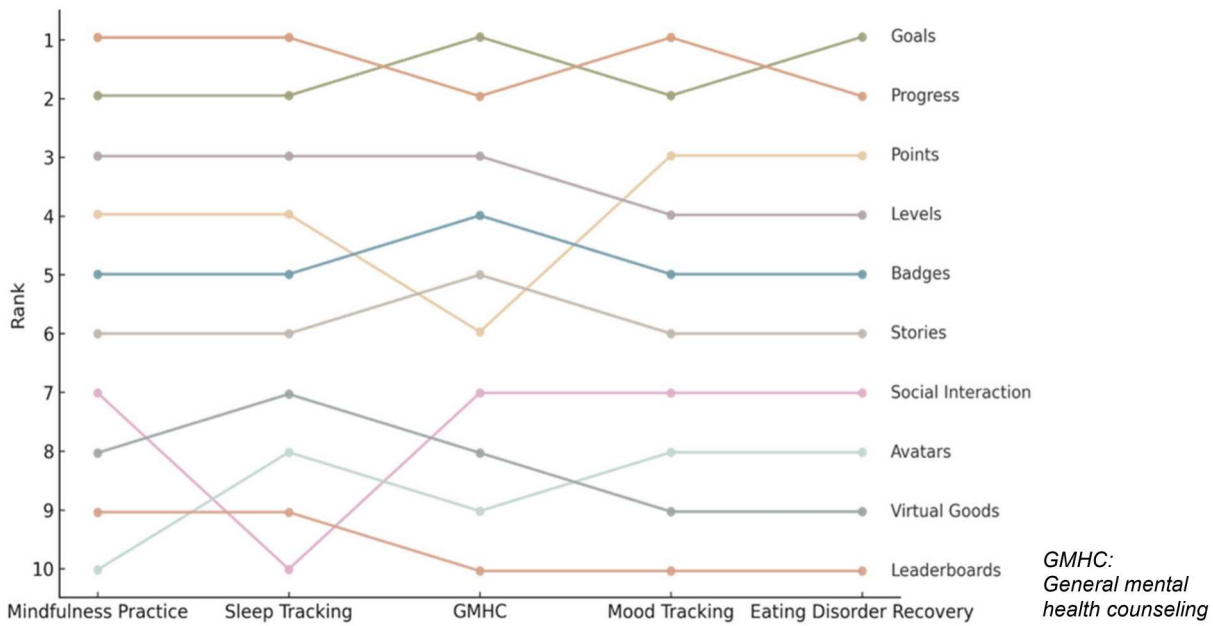


Figure 4. Ranking of game design elements across DMHI contexts.

directional clarity can ‘motivate [users] to continue because [they] know what they are trying to achieve’ (ID848, mindfulness practice). GDEs like *Badges* and

Levels supplement this by serving as intermediate way-points that break down a large goal into ‘reachable targets’ (ID623, mindfulness practice), making the overall

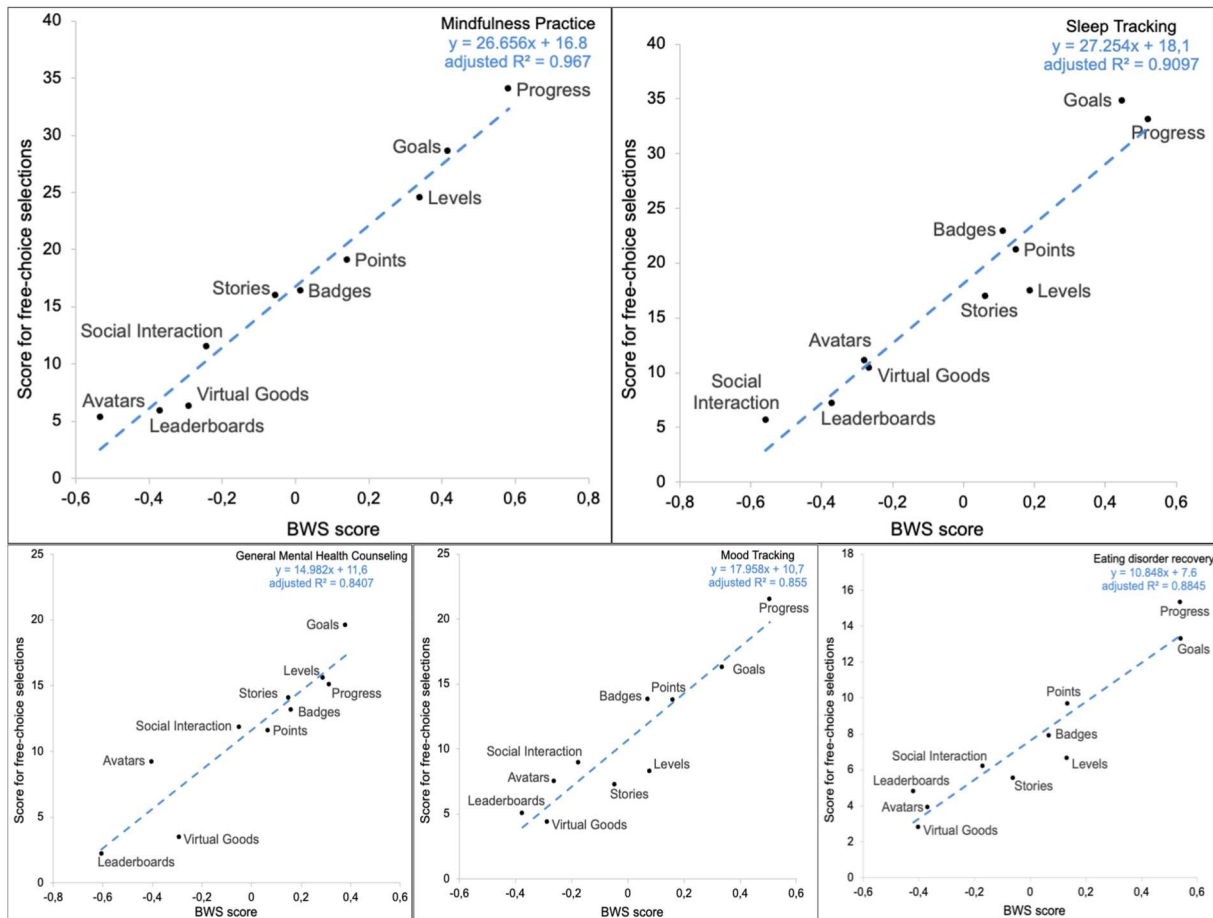


Figure 5. Relationship between BWS scores and free-choice selections across contexts.

Table 3. Results of conditional logit regression analyses for the five DMHI application contexts.

Mindfulness practice			Sleep tracking			General mental health counselling			Mood tracking			Eating disorder recovery		
GDE	Coef	SE	GDE	Coef	SE	GDE	Coef	SE	GDE	Coef	SE	GDE	Coef	SE
Progress	2.662	0.081	Progress	2.139	0.077	Goals	3.171	0.192	Progress	3.090	0.169	Goals	3.365	0.206
Goals	2.026	0.080	Goals	1.951	0.077	Progress	2.836	0.189	Goals	2.332	0.169	Progress	3.260	0.203
Levels	1.750	0.081	Levels	1.021	0.082	Levels	2.674	0.193	Points	1.347	0.170	Points	1.226	0.204
Points	1.156	0.086	Points	0.924	0.083	Badges	2.163	0.192	Levels	1.051	0.175	Levels	1.068	0.208
Badges	0.953	0.089	Badges	1.014	0.082	Stories	2.347	0.191	Badges	1.174	0.170	Badges	1.105	0.202
Stories	0.903	0.089	Stories	1.089	0.081	Points	1.833	0.194	Stories	0.967	0.172	Stories	0.663	0.207
SI	0.533	0.096	VG	0.464	0.090	SI	1.936	0.193	SI	1.076	0.173	SI	0.840	0.210
VG	0.547	0.096	Avatars	0.152	0.097	VG	0.469	0.215	Avatars	0.316	0.186	Avatars	−0.743	0.267
LB	0.248	0.102	LB	0.258	0.094	Avatars	0.187	0.225	VG	0.282	0.186	VG	−0.200	0.237
Avatars	0	-	SI	0	-	LB	0	-	LB	0	-	LB	0	-

SI: Social interaction; VG: Virtual goods; LB: Leaderboards.

Table 4. Results from the conditional logit regression with GDE × context interactions.

Interaction Term	Coef	SE	Pr(> z)	Interaction Term	Coef	SE	Pr(> z)
Levels x MP	1.279	0.344	< 0.001***	SI x ST	−1.893	0.288	< 0.001***
VG x MP	0.648	0.297	0.029*	VG x ST	0.605	0.286	0.034*
Avatars x MP	−0.523	0.349	0.133	Stories x ST	0.467	0.269	0.082
Stories x MP	0.399	0.284	0.160	Avatars x ST	0.427	0.320	0.182
SI x MP	−0.295	0.285	0.301	Goals x ST	−0.301	0.328	0.358
Progress x MP	0.361	0.370	0.330	Points x ST	0.175	0.300	0.559
Points x MP	0.203	0.311	0.513	Progress x ST	0.112	0.361	0.757
Goals x MP	0.104	0.348	0.764	Levels x ST	0.010	0.322	0.976
Badges x MP	−0.079	0.290	0.785	Badges x ST	0.006	0.278	0.982
Interaction Term	Coef	SE	Pr(> z)	Interaction Term	Coef	SE	Pr(> z)
Stories x GMHC	1.600	0.308	< 0.001***	Avatars x MT	0.953	0.330	0.004**
Levels x GMHC	1.367	0.357	< 0.001***	Goals x MT	−0.817	0.338	0.016*
Badges x GMHC	1.165	0.318	< 0.001***	VG x MT	0.565	0.307	0.066
VG x GMHC	0.933	0.328	0.004**	Levels x MT	−0.482	0.334	0.149
SI x GMHC	0.758	0.298	0.011*	Progress x MT	−0.376	0.365	0.302
Avatars x GMHC	0.878	0.354	0.013*	Points x MT	0.322	0.321	0.316
Points x GMHC	0.543	0.328	0.099	Stories x MT	0.186	0.288	0.518
Goals x GMHC	0.320	0.357	0.370	SI x MT	−0.105	0.282	0.710
Progress x GMHC	−0.273	0.368	0.457	Badges x MT	−0.042	0.298	0.888

MP: Mindfulness practice; ST: Sleep tracking; GMHC: General mental health counselling; MT: Mood tracking.
VG: Virtual Goods; SI: Social interaction; * $p < .05$; ** $p < .01$; *** $p < .001$.**Table 5.** Rationales for users' game design element preferences in DMHIs.

Category	Rationale	Main mechanisms	Exemplary GDEs
Function-related rationales	Meaningful orientation	<i>Direction-setting</i> : GDEs establish a clear forward-looking journey to foster a sense of purpose.	Goals, Badges, Levels
	Opportune feedback	<i>Path-reflection</i> : GDEs enable users to reflect on their journey over time. <i>Effort acknowledgement</i> : GDEs provide instant confirmation of task completion to create a sense of accomplishment.	Progress Points, Badges
	External support	<i>Providing rewards</i> : GDEs deliver tangible or symbolic rewards to enhance motivation and enjoyment. <i>Connection for belonging</i> : GDEs create a sense of community for emotional reassurance. <i>Exchange for informational support</i> : GDEs allow users to gain practical advice and new perspectives by exchanging experiences.	Virtual goods, Badges Social Interaction Social Interaction
Intervention-related rationales	Instrumental alignment	<i>Social influence for motivation</i> : GDEs motivate users through recognition by others. <i>Supportive utility</i> : GDEs are perceived as a useful tool that adds practical value to the intervention's core task.	Leaderboards Stories for GMHC Progress for MT
	Psychological safety	<i>Contextual appropriateness</i> : GDEs are not perceived as distracting or incongruent with the intervention's primary purpose. <i>Stress-free engagement</i> : GDEs avoid competition or pressure to maintain a calm atmosphere.	Badges, Stories
		<i>Non-judgmental, self-focused experience</i> : GDEs enable users to concentrate on personal growth without external judgment or benchmarks.	Progress
User-related rationales	Subjective appeal	<i>Alignment with personal interests</i> : GDEs resonate with a user's personal interests or social tendencies.	Stories, Social Interaction
	Experience mindset	<i>Aesthetic pleasure</i> : GDEs appeal through visual design and style. <i>Familiar and reliable engagement</i> : GDEs are chosen when past experience has proven them effective and reliable.	Badges, Avatars Goals, Progress, Badges
		<i>Novel and new experience</i> : GDEs are valued when they offer a new and different way to engage users.	Stories, Avatars

ST: Sleep tracking; GMHC: General mental health counselling; MT: Mood tracking.

journey ‘achievable without being pressured’ (ID1443, sleep tracking). In terms of path-reflection, users favour GDEs that enable them to reflect on their journey over time. By making the path and distance-to-go visible at a glance, such GDEs help users understand their current standing and next steps, thereby ‘stay[ing] motivated and engaged’ (ID742, mindfulness). *Progress* was most valued for this function, as it ‘provides an objective opportunity for self-reflection’ (ID216, general mental health counselling) through ‘a clear track of things to do and already done’ (ID164, eating disorder recovery).

Opportune Feedback. This rationale concerns user preferences for immediate, interpretable acknowledgment of their specific actions and efforts. This type of feedback focuses on timely validation that creates a satisfying, short-term reinforcement loop and impacts user preference in two main ways: (1) by acknowledging effort to create a sense of accomplishment, and (2) by providing rewards to enhance motivation. First, users favour GDEs that clearly confirm the successful completion of a task, which fosters a sense of accomplishment. For example, some users appreciate GDEs like *Points* and *Badges* that immediately ‘show [them] a result’ (ID187, general mental health counselling), which ‘gives a feeling of accomplishment’ (ID100, mood tracking) and evokes ‘the dopamine effect and to hook the user on the app’ (ID667, general mental health counselling). Second, users are motivated by GDEs that provide tangible or emotional rewards for their actions. Such rewards are ‘intuitive and easy to understand’ (ID350, sleep tracking), thereby going beyond simple acknowledgment to make the experience more enjoyable and compelling, without overwhelming users. To this end, *Points* are valued because they act as ‘an immediate reward’ (ID383, eating disorder recovery) and ‘convert [users’] efforts into something [they] can see and use’ (ID1523, mood tracking). Similar positive incentives can also come from *Badges* and *Virtual Goods*, which present a straightforward ‘reason for users to continue’ (ID1040, mindfulness).

External Support. Users prefer GDEs that connect them to sources of support beyond the self, such as peers who can provide encouragement, advice, and recognition. Preferences driven by this rationale stem from the intrinsic social function of *Social Interaction* and *Leaderboards*, which leverage community to enhance the intervention experience. This rationale manifests in three ways: (1) fostering belonging through connection, (2) informational support through experience exchange, and (3) driving motivation through social influence. First, users value *Social Interactions* that create a sense of community for emotional reassurance. Being able to ‘communicate with real people with similar problems’ (ID423, general mental health

counselling) helps users feel they are not alone and provides a crucial sense of being ‘heard, seen, and understood’ (ID383, eating disorder recovery). Beyond this emotional validation, the social connection is also a practical resource. By exchanging experiences, users can ‘reflect well on their own and [...] get another point of view and tips from the other person’ (ID390, general mental health counselling). The combination of emotional and informational support empowers users, reinforcing their commitment to their intervention journey. Lastly, social influence can motivate users, but it must be managed carefully to avoid causing stress. For some, the ability to compare with others is a straightforward way ‘to do more [...] than other people’ (ID794, mood tracking), turning a solitary activity into a more engaging social experience. However, many users caution that competition can also create pressure, which contradicts the fundamental purpose of DMHIs. Therefore, social motivation must be balanced against the risk of creating this pressure, focusing on being encouraging rather than overtly competitive.

4.3.2. Intervention-Related Rationales

Rationales in this category concern the particularities of a DMHI. Here, user preferences are determined by the perceived fit between the GDE and the DMHI’s overarching instrumental task or targeted behaviours (e.g. regular participation in mindfulness sessions, improved sleep quality). While the specific instrumental tasks differed across the five application contexts in our study, the underlying requirements for GDEs to be purposeful and non-disruptive to the user’s mental state remained constant. We identified two key rationales reflecting these contextual particularities: instrumental alignment and psychological safety.

Instrumental Alignment. This rationale reflects the user’s need for GDEs to (1) add practical value to the intervention task (supportive utility) and (2) remain aligned with the purpose of the DMHI (contextual appropriateness). On the one hand, users consistently favoured elements they perceived as functional tools that support them in performing the central task of the intervention, especially when a GDE provided a capability that the core intervention could not inherently offer. For example, in the context of mood tracking, one participant asserted that because ‘mood is something [people] can’t really count, looking at the progress helps them know how their emotion is’ (ID1543). Thus, *Progress* was highly valued in this context for making the abstract mood tangible. Similarly, in the general mental health context where users can find it challenging to articulate their emotional issues, *Stories* were preferred because they ‘help to express problems

better' (ID231), thereby facilitating a core requirement of the intervention. On the other hand, participants drew clear boundaries against GDEs that are incongruent with the DMHI's core purpose. For sleep tracking, participants repeatedly rejected social or competitive elements as inappropriate for an intimate and personal routine, with one stating that 'more is not needed, it may end up in a social competition' (ID1003). This sentiment was echoed by participants in mindfulness practice, who valued 'playful but non-distracting elements', cautioning that incongruent features 'could harm the productivity of the app' (ID559, ID814).

Psychological Safety. This rationale stems from the user's need for the DMHI to function as a psychologically safe space, where self-improvement can be pursued without fear of judgment or failure. Two mechanisms drive this rationale. First, participants prioritise a stress-free engagement with the intervention, which is primarily achieved by removing competition and pressure. The most consistent theme was the rejection of elements that 'put the user under pressure to perform well, which leads to more stress' (ID1308, sleep tracking). Participants explicitly warned against both social pressure (i.e. 'compare and compete with others', ID953, mood tracking) and design-induced pressure, such as punitive design like 'dropping levels if you break your strike' (ID1145, sleep tracking). Accordingly, the ideal is a calm atmosphere, where reinforcement and motivation feel gentle, such as 'giv[ing] a sense of achievement without being under too much pressure' (ID1485, sleep tracking:), or 'encourag[ing] the progression in a non-competitive way' (ID708, mood tracking). Second, participants appreciate a non-judgmental, self-focused experience that helps them to center on the individual's unique journey. For example, GDEs are valued when they provide 'an objective opportunity for self-reflection and the focus is on oneself' (ID216, general mental health counselling). This focus is especially crucial for people in a vulnerable state. Consequently, participants consistently prefer GDEs that 'primarily focus on the individual growth' (ID554, mindfulness practice), rather than being assessed against a standardised or competitive benchmark.

4.3.3. User-Related Rationales

Rationales in this group are rooted in individuals' personal characteristics, such as their aesthetic tastes, prior experiences, and intrinsic motivations. While function-related rationales address general user value, these rationales highlight that even a perfectly functional and appropriate GDE might be rejected if it fails to resonate with a user on a personal level. Our analysis identified two key rationales that capture

these individual differences: experience mindset and subjective appeal.

Experience Mindset. This rationale describes how a user's past experiences shape their current mindset, which in turn influences their preference for either familiar or novel GDEs. On the one hand, individuals' personal history with GDEs creates a mental model of what works for them. This mindset leads to a preference for familiar GDEs that have proven effective for them in the past, as this provides a reliable and low-effort path to motivation. For instance, two participants clearly articulated the preference for certain GDEs that 'also motivate [them] the most in other apps' (ID1194, mindfulness; ID346, sleep tracking). This explanation was substantiated by Chi-Square tests, which assess whether familiarity and preferences are correlated. The results indicate that participants were more likely to select GDEs they were familiar with (mindfulness practice: $\chi^2(1, 168) = 51.118$; sleep tracking: $\chi^2(1, 181) = 26.212$; general mental health counselling: $\chi^2(1, 116) = 24.101$; mood tracking: $\chi^2(1, 107) = 42.108$; eating disorder recovery: $\chi^2(1, 76) = 25.394$; all $p < .001$). On the other hand, an experience mindset can also drive individuals to seek novelty and new experiences, which may offer something different or more engaging. For instance, some users preferred the *Stories* because they are 'different from what [they] are used to' (ID636, mindfulness) and 'gives [them] a different way to engage' (ID378, general mental health counselling). This duality underscores that the influence of past experiences on users' GDE preference may not be universal.

Subjective Appeal. This rationale captures user preferences driven by a user's subjective, personal tastes. These choices are often rooted in a desire for (1) a more personally engaging experience or (2) an aesthetically pleasing experience. Users often project their personal interests and social tendencies onto their choice of GDEs. For instance, participants who enjoy reading are naturally drawn to *Stories*. Similarly, socially inclined users welcome chances to share and check in through *Social Interaction*. While many participants reject competition, a small group who are personally motivated by social comparison may appreciate a *Leaderboard*, 'as it is an incentive to keep up with others' (ID389, eating disorder recovery). Beyond personal interests, participants also emphasised the aesthetic qualities of GDEs. Their accounts reveal two complementary dimensions of aesthetics (Lavie and Tractinsky 2004). First, elements were valued for qualities of classical aesthetics, such as clarity, pleasantness, and harmony (Lavie and Tractinsky 2004), for example, when Avatars or Badges were chosen as they are 'cute and eye-catching' or 'aesthetically well-

designed' (ID346, sleep tracking; ID1454, mood tracking). Second, participants also appreciated aspects of expressive aesthetics, which highlight creativity and originality in design, making elements feel distinctive (Lavie and Tractinsky 2004). While the desire for expressive aesthetics may seem related to the preference for novelty in the Experience Mindset rationale, the two are conceptually distinct. The former reflects a user's evaluation of the design's intrinsic creative qualities (an artifact-centric judgment), while the latter is based on the design's novelty relative to the user's personal history of use (a user-centric, comparative judgment). For example, an experienced gamer might be attracted to *Stories* because of their unique and creative design (due to their expressive aesthetics). In contrast, a user with no prior experience with narrative-driven systems might be drawn to this GDE simply because it offers a completely new form of interaction, even if its design is not particularly original.

5. Discussion

This study investigates user preferences for GDEs in the domain of DMHIs and uncovers the underlying rationales for these preferences. The following sections synthesise three key insights of this study and contrast them with findings from prior work (see Table 6). Afterward, we discuss the theoretical and practical implications, reflect on the study's limitations, and identify directions for future research.

Table 6. Key insights of this study compared to prior research.

Prior research	Findings of this study
Gamification research has emphasised that the effects of similar gamification designs vary substantially across application contexts, considering context as a strong influential factor for user preferences (Klock et al. 2020; Koivisto and Hamari 2019).	GDE preferences remain remarkably stable across five different DMHI application contexts. Contextual differences emerge selectively for specific GDEs rather than uniformly across all elements.
Individual GDEs have often been characterised as uniformly beneficial or detrimental, and mixed or heterogeneous effects have been reported with limited understanding of why the same GDE can evoke both positive and negative perceptions (e.g. Almeida et al. 2023).	Seven rationales explain users' GDE preferences in DMHIs. Some rationales were used by users both to justify and to reject the same GDE within a single context, highlighting the dual evaluative potential of these rationales for GDE preferences.
Task–technology fit conceptualises fit as a single correspondence among task, technology, and individual (Goodhue and Thompson 1995). Gamification research on tailored design has focused primarily on matching elements to user characteristics (Klock et al. 2020).	GDE preferences emerge through layered alignment between GDE functionality, intervention characteristics, and user characteristics, leading to the concepts of element–individual fit and element–intervention fit.

5.1. General patterns and contextual nuances of GDE preferences in DMHIs

A primary finding of our study is the notable consistency in GDE preference rankings across all five DMHI contexts, despite the variation in intervention characteristics. These rankings also strongly resemble the findings of previous BWS studies on users' GDE preferences (Berger and Jung 2024; Schmidt-Kraepelin et al. 2019; Schöbel, Söllner, and Leimeister 2016), especially in health-related domains. A plausible explanation for these strong similarities is that, within health-oriented applications, users generally value GDEs that help them focus on their individual journey and self-development toward a desired outcome, such as a healthier lifestyle (Schmidt-Kraepelin et al. 2019). Another example of general patterns is the comparatively higher acceptance of *Stories* in DMHIs relative to other domains such as physical activity, where competition-oriented *Leaderboards* are often preferred. This finding suggests that narrative mechanisms may be especially suited to the nature of DMHIs and opens pathways for more nuanced research on how *Stories* can be meaningfully adapted to the unique requirements of DMHIs. Such work would also respond to recent calls in gamification research to move beyond structural gamification, which mainly relies on *Points*, *Badges*, and *Leaderboards* as a detachable layer on top of a task, toward content gamification, which integrates gameful features into the core task to create more immersive and engaging experiences (Koivisto and Hamari 2019).

At the same time, our results reveal specific deviations in preferences for certain GDEs across DMHI application contexts. Such deviations were not random but concentrated in elements whose alignment with the intervention's primary purpose and targeted behaviours varies by context. For instance, the rejection of *Social Interaction* was far more pronounced in the sleep tracking context, where the element's inherent function of encouraging connectivity conflicts with the intervention's primary goal of self-monitoring and its targeted behaviours of achieving a calm state conducive to rest. Another illustration of the nuanced deviations is that *Avatars* were ranked low in mindfulness practice, but received more attention in mood tracking, where users consider them supportive for self-expression. This pattern may reflect the stronger emphasis on repeated emotional self-reflection and awareness in this context. These nuances underscore the complexity of gamification in this sensitive domain and highlight the need to understand boundaries of instrumental alignment. Especially, whereas many GDEs like *Points*

build on a relatively uniform design concept, *Avatars* encompass a wide design space. It can range from a simple icon to a highly customizable, realistic representation, with each design choice likely affording a different user experience (Miao et al. 2022). Thus, future studies could investigate specific design characteristics and contextual factors that determine whether an avatar is perceived as a valuable tool in DMHIs.

5.2. Duality in rationales for GDE preferences

Another interesting insight from our results is the inherent duality of certain rationales underlying GDE preferences. Specifically, the influence of certain rationales on user preferences may not be universal. Instead, the same rationale may be perceived as either supportive or counterproductive, depending on the user's mindset, prior experiences, or situational needs. A clear example of this duality emerges in the rationale of Experience Mindset. Familiarity with certain GDEs can enhance trust, reduce learning costs, and increase motivation, as users perceive these elements as reliable and proven (Rodrigues et al. 2022). Yet this reliance on familiar mechanisms may also undermine engagement over time by reducing novelty (Rodrigues et al. 2022). Conversely, while novel GDEs like *Avatars* may attract interest by offering fresh ways of engaging, they may also create confusion or resistance among users who feel overwhelmed by unfamiliar designs. Thus, prior experience can shape both positive and negative attitudes toward the same element. A similar duality is shown in the social elements, such as *Social Interaction* and *Leaderboards*. These GDEs are valued by a minority of users whose Subjective Appeal leans toward competition or social activities. However, these same elements are strongly rejected by the majority, who experience them as intrusive or pressuring, leading to counterproductive comparisons and feelings of being judged.

This duality in rationales helps explain why some GDEs (e.g. *Avatars* or *Social Interaction*) elicit sharply divergent evaluations across users, even within the same DMHI context. These findings also align with extant research that challenges a simplistic view of GDEs as universally 'good' or 'bad' (e.g. Almeida et al. 2023). Even within a single DMHI, there is no one-size-fits-all answer to the appropriateness of certain GDEs.

5.3. A multi-level fit perspective for GDE preference in DMHIs

The categorisation of our seven rationales aligns well with the task-(individual-) technology fit (TTF) theory and its extension, the technology-to-performance

chain (TPC) model (Goodhue and Thompson 1995). TTF posits that the impact of a technology depends on its 'fit' with users and the task at hand (Goodhue and Thompson 1995). Specifically, TTF is defined as the correspondence between task requirements, the technology's functionality, and the individual's own characteristics, while TPC conceptualises TTF as an antecedent to user attitudes that in turn shape utilisation behaviour (Goodhue and Thompson 1995).

Our findings provide a contextualised extension of the TTF in the domain of gamification for DMHIs. While function-related rationales capture how the intrinsic functionalities of a GDE (the 'technology') afford the actions required by users, user-related rationales address user characteristics, like 'individual' interests and capabilities, and intervention-related rationales reflect how well those functionalities fit the 'task' and context of the DMHI. Viewed through this theoretical lens, the first two categories of rationales capture the alignment between the characteristics of a GDE and users' individual needs, constituting *element-individual fit*. The intervention-related rationales, however, point to *element-intervention fit*, referring to the alignment between the characteristics of a GDE, especially its functionality, and the requirements of the intervention context. Our findings suggest that both forms of fit contribute positively to the formation of users' GDE preferences. Combined with the discussion about general patterns, contextual deviations, and preference polarisation in the above two sections, we propose a layered view of fit for explaining rationales for users' GDE preferences in DMHIs (see Figure 6). The model further highlights the importance of understanding user preferences by considering their role in shaping user behaviours (Hertzum 2025), such as the acceptance and continued use of gamified DMHIs.

Given the strong general patterns in GDE preferences, we argue that a large part of preference is formed by function-related rationales, which constitute a broad baseline for element-individual fit, whereas user-related rationales capture how individual characteristics condition this fit. Intervention-related rationales, which explain deviations in preferences across contexts, add a second layer of *element-intervention fit*. This layer plays a crucial but more selective role, proving decisive only for specific GDEs in specific contexts. This layered perspective helps explain why some GDEs, such as *Goals* and *Progress*, appear almost universally valued (i.e. they satisfy requirements of both layers), while others, like *Avatars* or *Social Interaction*, are more contingent on context and individual differences.

Taken together, our layered model of fit suggests these two 'fits' are not equally weighted, which

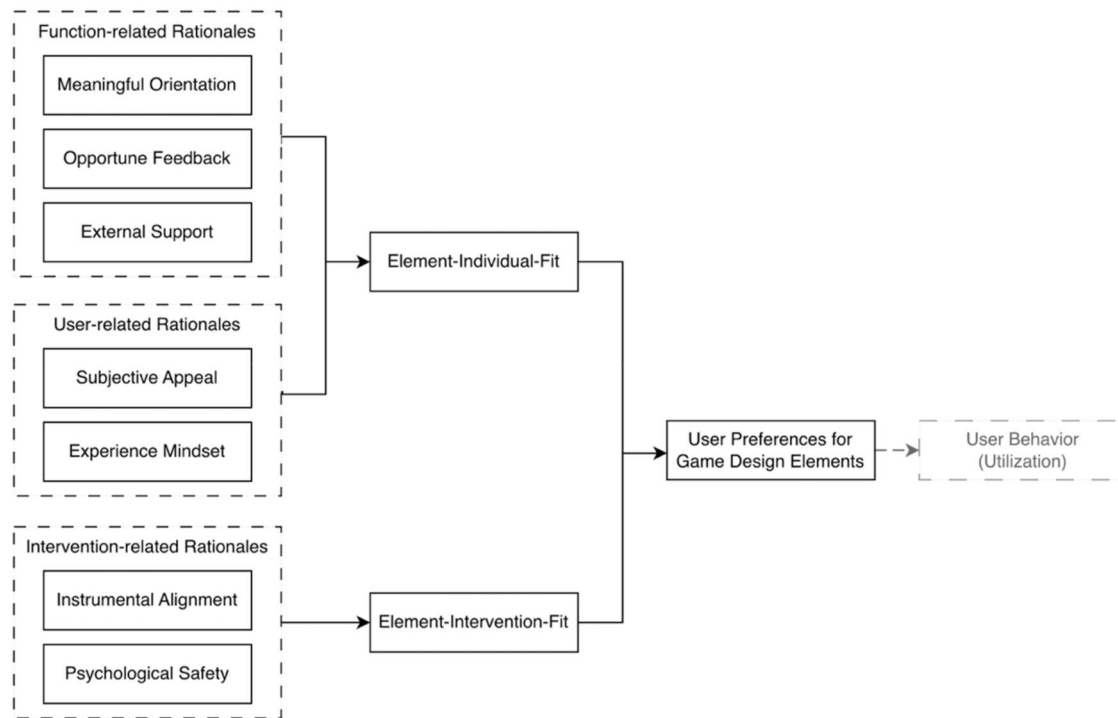


Figure 6. A layered fit model of user rationales for GDE preferences in DMHIs.

challenges prior research that often emphasised the high context-sensitivity of gamification. Accordingly, this view highlights the need to revisit theoretical assumptions about context sensitivity and to explore how different layers of fit interact in shaping user preferences. Furthermore, our model also aligns with recent calls in gamification research for moving beyond a sole focus on user characteristics toward a broader conception of tailoring that incorporates task and contextual factors (Klock et al. 2020). Our findings provide empirical support for this direction: in DMHIs, effective gamification design cannot be reduced to matching elements with static user profiles but requires accounting for the layered alignment of technology, task, and individual. Future research should therefore examine how multi-level tailoring strategies can be operationalised in DMHIs.

5.4. Implications

Based on the principal findings discussed above, our study yields several theoretical and practical implications for research and design in gamified DMHIs (see Table 7).

From a research perspective, the finding of stable general patterns with selective contextual deviations challenges common assumptions about the strong context-sensitivity of gamification. Prior work often emphasises that the effects of similar gamification

designs may vary substantially across application contexts (Klock et al. 2020; Koivisto and Hamari 2019). While the contextual deviations observed in comparison with extant findings on user preferences (e.g. for *Stories* and *Leaderboards*) confirm that context matters, they indicate that its influence is selective rather than pervasive. By clarifying that context does not uniformly reshape overall preference structures but instead moderates particular GDE preferences, our study refines the understanding of context effects in gamified IS.

In addition, the finding of certain rationales' dual effects on user preferences provides an explanation for the mixed effects of gamification reported in prior research (Cheng et al. 2019; Koivisto and Hamari 2019). The same underlying rationale can be interpreted in opposing ways, such that it may support or undermine preference depending on how users interpret it. This perspective helps explain why ambivalent responses to GDEs can occur even within the same DMHI context and complements prior work on heterogeneous effects of a specific GDE (e.g. Fallon et al. 2020). Rather than assuming that a given GDE is either beneficial or detrimental, our study shows that the rationales underlying users' evaluations can operate in both supportive and conflicting directions. This insight paves the way for more fine-grained research investigating the boundary conditions under which specific GDEs enhance or undermine their intended effects in DMHIs.

Table 7. Implications of the study.

Principal findings	Theoretical implications	Practical implications
Stable general patterns with selective contextual deviations	- Refines prior assumptions about the high context-sensitivity of gamification by showing that context functions as a selective moderator of specific design rather than as a dominant determinant of overall preference structures - Distinguishes between GDEs with stable cross-context preference hierarchies and GDEs subject to context-specific shifts, thereby clarifying where context effects are likely to matter in gamified IS	- Provides guidance for distinguishing between GDEs with stable cross-context acceptance and those showing context-dependent variation - Suggests directing the contextual design effort toward GDEs with context-dependent preferences
Dual evaluative potential of rationales	- Demonstrates that rationales underlying GDE preferences can carry dual evaluative potential, such that the same rationale may support or undermine preference depending on how it is interpreted by users - Provides a structured explanation of ambivalent responses to GDEs in DMHIs and establishes a foundation for a more in-depth investigation of the boundary conditions for the effectiveness of GDEs in DMHIs	Indicates that GDEs likely to activate opposing interpretations among users are better offered as optional rather than mandatory components in DMHIs
Layered alignment underlying GDE preference formation	- Shows that users' GDE preference formation in DMHIs reflects simultaneous alignment at functional, contextual, and individual levels, rather than a single dominant source of fit - Contextualises the TTF perspective to gamified DMHIs by distinguishing element-individual fit and element-intervention fit and showing how these jointly shape users' attitudes toward GDEs	Encourages DMHI designers to move beyond evaluating gamification primarily in terms of engagement effects, and to consider how GDEs align functionally with the intervention task, contextually with the therapeutic logic, and individually with users' characteristics

Lastly, the layered alignment underlying GDE preference formation contextualises the TTF for the mental health domain and offers a structured explanatory mechanism for understanding how gamification operates in

mental health settings. While TTF conceptualises fit as the correspondence among task requirements, technological characteristics, and individual attributes (Goodhue and Thompson 1995), our results indicate that in DMHIs, fit is layered and unequally weighted, emerging through two distinct forms: *element-individual fit* and *element-intervention fit*. This layered view may help account for both stable cross-context preferences (e.g. *Progress* and *Goals* via strong element-individual fit) and selective contextual sensitivities (e.g. low preference for *Leaderboards* via element-intervention misfit), while also providing a basis for understanding ambivalence in user evaluations. The layered framework forms a theoretical basis for future empirical studies that seek to model how the identified rationales interact in shaping user evaluations and outcomes. In this sense, it contributes not only descriptive insight into preference formation, but also a fundamental structure to guide subsequent explanatory and causal research on gamified DMHIs.

From a practical perspective, these findings provide guidance on how user preferences can inform gamification design in DMHIs. The presence of stable cross-context preference patterns suggests that not all GDEs require extensive contextual redesign, whereas GDEs whose preference patterns vary across settings warrant greater contextual attention in design decisions. In addition, because some GDEs are likely to activate opposing interpretations among users, they are better offered as optional rather than mandatory components in DMHIs. More broadly, evaluating gamification in mental health contexts benefits from going beyond engagement metrics alone and considering whether individual GDEs align with the intervention task, the therapeutic logic of the setting, and relevant user characteristics.

5.5. Limitations and future research

This study is not without limitations, which at the same time provide valuable avenues for future research. First, participants were introduced to GDEs through mockups and descriptions rather than within a real functioning system. Although user preference serves as an appropriate antecedent for user behaviour (Goodhue and Thompson 1995), users' actual behaviour may deviate from self-reported preferences. In this regard, it would also be worthwhile to validate whether users' self-reported GDE preferences and actual engagement with GDEs turn out to be consistent or not. Second, our study focused on app-based DMHIs, which represent only one of several technological enablers through which GDEs can be delivered. Gamification in this domain can also be realised through more immersive or embodied digital formats, such as extended reality

or wearable-based interventions, which differ in their modes of interaction and the experiences they can support. These differences may shape which GDEs users prefer and why. Future research could examine whether the preference rationales identified in this study and the proposed layered fit perspective remain applicable across different technology enablers. Third, each GDE was represented by a single visualisation and description, whereas in practice, these elements allow for diverse design variations. The chosen representation likely influenced participants' evaluations. Investigating different design variants of the same GDE could yield deeper insights into why some elements are less preferred and how their design could be improved. Fourth, our analysis focused on GDEs in isolation, whereas in practice, these elements are often implemented as bundles or combinations. Synergistic or conflicting effects can emerge from the interplay between GDEs. Future work could investigate GDE combinations to understand how they collectively shape user perception. Fifth, our analysis of rationales relied on responses to a single open-ended survey question. Compared to existing BWS studies on users' GDE preferences, we are the first to take this step of capturing users' rationales for their preferences, and we complemented participants' qualitative statements with insights from statistical analyses of the BWS data. Despite the value of these initial insights, future research could build on this work by employing more extensive qualitative approaches, such as in-depth interviews or expert panels, to explore these mechanisms in greater depth. Such qualitative exploration could also inform subsequent larger-scale empirical investigations that further examine how and under which conditions specific GDEs influence user engagement in DMHIs. Finally, this study focuses on understanding the rationales underlying users' preferences for GDEs in the context of DMHIs and, therefore, considers a broad user population rather than focusing on specific subgroups. While the study shows that user-related factors influence the formation of preferences, it does not explicitly examine how preferences may differ across particular demographic, clinical, or cultural groups. The conceptual insights derived from the identified rationales and the proposed framework may provide a broadly applicable lens for understanding how user preferences for GDEs emerge across different user populations. However, future research focusing on specific user populations or applications in particular markets should carefully consider the potential limits to the generalizability of the empirical preference rankings reported in this study and may conduct more targeted investigations to examine how preferences for GDEs vary across different user segments.

6. Conclusion

Gamification has often been discussed as a promising strategy to sustain engagement in DMHIs. However, its design has largely relied on assumptions borrowed from other domains, such as education or marketing, with little research on its specific suitability for this sensitive domain. By examining user preferences across five DMHI application contexts and uncovering the rationales behind them, this study demonstrates that gamification in DMHIs follows its own logic. Preferences here are neither uniform nor easily transferable but emerge from a layered interplay between what elements afford, what intervention contexts demand, and what users themselves value. This perspective challenges the tendency to view GDEs as universally effective and calls instead for approaches that embrace ambivalence and accommodate diversity. By doing so, the field can help bridge the critical gap between the vast potential of DMHIs and their real-world impact, ensuring that the interactive technologies we build are not only engaging but genuinely supportive of well-being.

Author contributions

CRedit: **Shanshan Hu:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing; **Manuel Schmidt-Kraepelin:** Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing; **Simon Krohmann:** Methodology, Validation, Writing – original draft, Writing – review & editing; **Scott Thiebes:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing; **Ali Sunyaev:** Resources, Supervision, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Almeida, C., M. Kalinowski, A. Uchôa, and B. Feijó. 2023. "Negative Effects of Gamification in Education Software: Systematic Mapping and Practitioner Perceptions." *Information and Software Technology* 156:107142. <https://doi.org/10.1016/j.infsof.2022.107142>.
- Andrews, G., A. Basu, P. Cuijpers, M. Craske, P. McEvoy, C. English, and J. Newby. 2018. "Computer Therapy for the Anxiety and Depression Disorders Is Effective, Acceptable and Practical Health Care: An Updated Meta-analysis." *Journal of Anxiety Disorders* 55 (1): 70–78. <https://doi.org/10.1016/j.janxdis.2018.01.001>.
- Baron, K.G., J. Duffecy, K. Reid, M. Begale, and L. Caccamo. 2018. "Technology-Assisted Behavioral Intervention to Extend Sleep Duration: Development and Design of the

- Sleep Bunny Mobile App.” *JMIR Mental Health* 5 (1): e3. <https://doi.org/10.2196/mental.8634>.
- Berger, M., and C. Jung. 2024. “Gamification Preferences in Nutrition Apps: Toward Healthier Diets and Food Choices.” *Digital Health* 10:20552076241260482. <https://doi.org/10.1177/20552076241260482>.
- Berger, M., S. Lahmer, M. Reuther, and M. Schoch. 2023. “Preferred Gamification Elements in a Health Behavior Change Support System for Stress Management.” Proceedings of the 56th Hawaii international conference on system sciences.
- Braun, V., and V. Clarke. 2006. “Using Thematic Analysis in Psychology.” *Qualitative Research in Psychology* 3 (2): 77–101. <https://doi.org/10.1191/1478088706qp063oa>.
- Brown, M., N. O’Neill, H. van Woerden, P. Eslambolchilar, M. Jones, and A. John. 2016. “Gamification and Adherence to Web-Based Mental Health Interventions: A Systematic Review.” *JMIR Mental Health* 3 (3): e39. <https://doi.org/10.2196/mental.5710>.
- Burbach, F., F. Lecce, V. M. Allen, and C. M. Porter. 2022. “A Conversational, Virtual, Avatar-led Cognitive Behavioral Therapy app Intervention for Improving the Quality of Life and Mental Health of People with Epilepsy: Protocol for a Randomized Controlled Trial.” *JMIR Research Protocols* 11 (11): e40261. <https://doi.org/10.2196/40261>.
- Cheng, V. W. S. 2020. “Recommendations for Implementing Gamification for Mental Health and Wellbeing.” *Frontiers in Psychology* 11:586379. <https://doi.org/10.3389/fpsyg.2020.586379>.
- Cheng, V. W. S., T. Davenport, D. Johnson, K. Vella, and I. B. Hickie. 2019. “Gamification in Apps and Technologies for Improving Mental Health and Well-Being: Systematic Review.” *JMIR Mental Health* 6 (6): e13717. <https://doi.org/10.2196/13717>.
- Cheng, V. W. S., T. A. Davenport, D. Johnson, K. Vella, J. Mitchell, and I. B. Hickie. 2018. “An app That Incorporates Gamification, Mini-games, and Social Connection to Improve Men’s Mental Health and Well-Being (MindMax): Participatory Design Process.” *JMIR Mental Health* 5 (4): e11068. <https://doi.org/10.2196/11068>.
- Deterding, S., D. Dixon, R. Khaled, and L. E. Nacke. 2011. “From Game Design Elements to Gamefulness: Defining “Gamification”.” Proceedings of the 15th international academic MindTrek conference: envisioning future media environments, Tampere, Finland.
- Etikan, I., S. A. Musa, and R. S. Alkassim. 2016. “Comparison of Convenience Sampling and Purposive Sampling.” *American Journal of Theoretical and Applied Statistics* 5 (1): 1–4. <https://doi.org/10.11648/j.ajtas.20160501.11>.
- Fallon, M., M. Schmidt-Kraepelin, S. Thiebes, S. L. Warsinsky, and A. Sunyaev. 2020. “Social Comparison in MHealth: The Role of Similar Others and Feelings of Envy.” Proceedings of the 41st international conference on information systems (ICIS), Hyderabad, India.
- Firth, J., J. Torous, J. Nicholas, R. Carney, A. Prapat, S. Rosenbaum, and J. Sarris. 2017. “The Efficacy of Smartphone-based Mental Health Interventions for Depressive Symptoms: A Meta-analysis of Randomized Controlled Trials.” *World Psychiatry* 16 (3): 287–298. <https://doi.org/10.1002/wps.v16.3>.
- Fitzgerald, M., and G. Ratcliffe. 2020. “Serious Games, Gamification, and Serious Mental Illness: A Scoping Review.” *Psychiatric Services* 71 (2): 170–183. <https://doi.org/10.1176/appi.ps.201800567>.
- Fitz-Walter, Z., D. Tjondronegoro, and P. Wyeth. 2011. “Orientation Passport: Using Gamification to Engage University Students.” Proceedings of the 23rd Australian computer-human interaction conference, Canberra, Australia.
- Fitz-Walter, Z., P. Wyeth, D. Tjondronegoro, and B. Scott-Parker. 2013. “Driven to Drive: Designing Gamification for a Learner Logbook Smartphone Application.” Proceedings of the first international conference on gameful design, research, and applications.
- Flynn, T. N., J. J. Louviere, T. J. Peters, and J. Coast. 2007. “Best–Worst Scaling: What It Can Do for Health Care Research and How to Do It.” *Journal of Health Economics* 26 (1): 171–189. <https://doi.org/10.1016/j.jhealeco.2006.04.002>.
- Gan, D. Z., L. McGillivray, J. Han, H. Christensen, and M. Torok. 2021. “Effect of Engagement with Digital Interventions on Mental Health Outcomes: A Systematic Review and Meta-analysis.” *Frontiers in Digital Health* 3 (1): 764079. <https://doi.org/10.3389/fdgh.2021.764079>.
- Goodhue, D. L., and R. L. Thompson. 1995. “Task-technology fit and Individual Performance.” *MIS Quarterly* 19:213–236. <https://doi.org/10.2307/249689>.
- Guo, S., W. Deng, H. Wang, J. Liu, X. Liu, X. Yang, C. He, Q. Zhang, B. Liu, and X. Dong. 2021. “The Efficacy of Internet-Based Cognitive Behavioural Therapy for Social Anxiety Disorder: A Systematic Review and Meta-Analysis.” *Clinical Psychology & Psychotherapy* 28 (3): 656–668. <https://doi.org/10.1002/cpp.2528>.
- Hamari, J. 2017. “Do Badges Increase User Activity? A Field Experiment on the Effects of Gamification.” *Computers in Human Behavior* 71:469–478. <https://doi.org/10.1016/j.chb.2015.03.036>.
- Hassan, L., J. Rantalainen, N. Xi, H. Pirkkalainen, and J. Hamari. 2020. “The Relationship between Player Types and Gamification Feature Preferences.” International GamiFIN conference.
- Hertzum, M. 2025. “Understanding Preference: A Meta-analysis of User Studies.” *International Journal of Human-Computer Studies* 195:103408. <https://doi.org/10.1016/j.ijhcs.2024.103408>.
- Hoffmann, A., C. A. Christmann, and G. Bleser. 2017. “Gamification in Stress Management Apps: A Critical app Review.” *JMIR Serious Games* 5 (2): e13. <https://doi.org/10.2196/games.7216>.
- Hopkin, G., H. Coole, F. Edelman, L. Ayiku, R. Branson, P. Campbell, S. Cooper, and M. Salmon. 2025. “Toward a new Conceptual Framework for Digital Mental Health Technologies: Scoping Review.” *JMIR Mental Health* 12: e63484. <https://doi.org/10.2196/63484>.
- Huotari, K., and J. Hamari. 2017. “A Definition for Gamification: Anchoring Gamification in the Service Marketing Literature.” *Electronic Markets* 27 (1): 21–31. <https://doi.org/10.1007/s12525-015-0212-z>.
- Jingili, N., S. S. Oyelere, M. B. Nyström, and L. Anyshchenko. 2023. “A Systematic Review on the Efficacy of Virtual Reality and Gamification Interventions for Managing Anxiety and Depression.” *Frontiers in Digital Health* 5:1239435. <https://doi.org/10.3389/fdgh.2023.1239435>.
- Johnson, D., S. Deterding, K.-A. Kuhn, A. Staneva, S. Stoyanov, and L. Hides. 2016. “Gamification for Health

- and Wellbeing: A Systematic Review of the Literature.” *Internet Interventions* 6 (1): 89–106. <https://doi.org/10.1016/j.invent.2016.10.002>.
- Kirchner-Krath, J., M. Altmeyer, L. Schürmann, B. Kordyaka, B. Morschheuser, A. C. T. Klock, L. Nacke, J. Hamari, and H. F. von Korflesch. 2024. “Uncovering the Theoretical Basis of User Types: An Empirical Analysis and Critical Discussion of User Typologies in Research on Tailored Gameful Design.” *International Journal of Human-Computer Studies* 190:103314. <https://doi.org/10.1016/j.ijhcs.2024.103314>.
- Klock, A. C. T., I. Gasparini, M. S. Pimenta, and J. Hamari. 2020. “Tailored Gamification: A Review of Literature.” *International Journal of Human-Computer Studies* 144 (102495): 102495. <https://doi.org/10.1016/j.ijhcs.2020.102495>.
- Koivisto, J., and J. Hamari. 2014. “Demographic Differences in Perceived Benefits from Gamification.” *Computers in Human Behavior* 35:179–188. <https://doi.org/10.1016/j.chb.2014.03.007>.
- Koivisto, J., and J. Hamari. 2019. “The Rise of Motivational Information Systems: A Review of Gamification Research.” *International Journal of Information Management* 45 (1): 191–210. <https://doi.org/10.1016/j.ijinfomgt.2018.10.013>.
- Kumar, A., M. Wang, A. Riehm, E. Yu, T. Smith, and A. Kaplin. 2020. “An Automated Mobile Mood Tracking Technology (Mood 24/7): Validation Study.” *JMIR Mental Health* 7 (5): e16237. <https://doi.org/10.2196/16237>.
- Lattie, E. G., C. Stiles-Shields, and A. K. Graham. 2022. “An Overview of and Recommendations for More Accessible Digital Mental Health Services.” *Nature Reviews Psychology* 1 (2): 87–100. <https://doi.org/10.1038/s44159-021-00003-1>.
- Lavie, T., and N. Tractinsky. 2004. “Assessing Dimensions of Perceived Visual Aesthetics of web Sites.” *International Journal of Human-Computer Studies* 60 (3): 269–298. <https://doi.org/10.1016/j.ijhcs.2003.09.002>.
- Lehtoranta, S., N. Xi, and J. Hamari. 2024. “Gamification and Employee Well-Being: A Systematic Literature Review.” Hawaii international conference on system sciences.
- Li, X., C. Lu, J. Peltonen, and Z. Zhang. 2019. A Statistical Analysis of Steam User Profiles Towards Personalized Gamification. GamiFIN.
- Liang, Z., E. Melcer, N. H. Hoang, and K. Khotchasing. 2024. “Sleep Hygiene Games and Gamification: Where Are We and Where Are We Heading?” *Frontiers in Sleep* 4:1–13. <https://doi.org/10.3389/frsle.2025.1607117>.
- Linardon, J., A. Shatte, M. Messer, J. Firth, and M. Fuller-Tyszkiewicz. 2020. “E-mental Health Interventions for the Treatment and Prevention of Eating Disorders: An Updated Systematic Review and Meta-analysis.” *Journal of Consulting and Clinical Psychology* 88 (11): 994–1007. <https://doi.org/10.1037/ccp0000575>.
- Litvin, S., R. Saunders, M. A. Maier, and S. Lüttke. 2020. “Gamification as an Approach to Improve Resilience and Reduce Attrition in Mobile Mental Health Interventions: A Randomized Controlled Trial.” *PLoS One* 15 (9): e0237220. <https://doi.org/10.1371/journal.pone.0237220>.
- Liu, D., R. Santhanam, and J. Webster. 2017. “Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems.” *MIS Quarterly* 41 (4): 1011–1034. <https://doi.org/10.25300/MISQ/2017/41.4.01>.
- Louviere, J., I. Lings, T. Islam, S. Gudergan, and T. Flynn. 2013. “An Introduction to the Application of (Case 1) Best–Worst Scaling in Marketing Research.” *International Journal of Research in Marketing* 30 (3): 292–303. <https://doi.org/10.1016/j.ijresmar.2012.10.002>.
- Lowry, P. B., T. Dinev, and R. Willison. 2017. “Why Security and Privacy Research Lies at the Centre of the Information Systems (IS) Artefact: Proposing a Bold Research Agenda.” *European Journal of Information Systems* 26 (6): 546–563. <https://doi.org/10.1057/s41303-017-0066-x>.
- McFadden, D. 1972. “Conditional Logit Analysis of Qualitative Choice Behavior.” In *Frontiers in Econometrics*, edited by P. Zarembka, 105–142. Academic Press.
- Mekler, E. D., F. Brühlmann, K. Opwis, and A. N. Tuch. 2013. “Disassembling Gamification: The Effects of Points and Meaning on User Motivation and Performance.” CHI ‘13 extended abstracts on human factors in computing systems, Paris, France.
- Miao, F., I. V. Kozlenkova, H. Wang, T. Xie, and R. W. Palmatier. 2022. “An Emerging Theory of Avatar Marketing.” *Journal of Marketing* 86 (1): 67–90. <https://doi.org/10.1177/0022242921996646>.
- Mohr, D. C., S. M. Schueller, E. Montague, M. N. Burns, and P. Rashidi. 2014. “The Behavioral Intervention Technology Model: An Integrated Conceptual and Technological Framework for EHealth and MHealth Interventions.” *Journal of Medical Internet Research* 16 (6): e146. <https://doi.org/10.2196/jmir.3077>.
- Morschheuser, B., L. Hassan, K. Werder, and J. Hamari. 2018. “How to Design Gamification? A Method for Engineering Gamified Software.” *Information and Software Technology* 95:219–237. <https://doi.org/10.1016/j.infsof.2017.10.015>.
- Nour, M. M., A. S. Rouf, and M. Allman-Farinelli. 2018. “Exploring Young Adult Perspectives on the use of Gamification and Social Media in a Smartphone Platform for Improving Vegetable Intake.” *Appetite* 120:547–556. <https://doi.org/10.1016/j.appet.2017.10.016>.
- Orme, B. K. 2009. “Anchored Scaling in MaxDiff Using Dual Response.” *Sawtooth Software. Research Paper Series*.
- Richards, D., and T. Richardson. 2012. “Computer-based Psychological Treatments for Depression: A Systematic Review and Meta-analysis.” *Clinical Psychology Review* 32 (4): 329–342. <https://doi.org/10.1016/j.cpr.2012.02.004>.
- Rodrigues, L., F. D. Pereira, A. M. Toda, P. T. Palomino, M. Pessoa, L. S. G. Carvalho, D. Fernandes, E. H. Oliveira, A. I. Cristea, and S. Isotani. 2022. “Gamification Suffers from the Novelty Effect but Benefits from the Familiarization Effect: Findings from a Longitudinal Study.” *International Journal of Educational Technology in Higher Education* 19 (1): 13. <https://doi.org/10.1186/s41239-021-00314-6>.
- Sardi, L., A. Idri, and J. L. Fernández-Alemán. 2017. “A Systematic Review of Gamification in e-Health.” *Journal of Biomedical Informatics* 71:31–48. <https://doi.org/10.1016/j.jbi.2017.05.011>.
- Schmidt-Kraepelin, M., S. Thiebes, S. Schöbel, and A. Sunyaev. 2019. “Users’ Game Design Element Preferences in Health Behavior Change Support Systems for Physical Activity: A Best–Worst–Scaling Approach.” Proceedings of the 40th international conference on information systems (ICIS), Munich, Germany.

- Schöbel, S., M. Söllner, and J. M. Leimeister. 2016. "The Agony of Choice—Analyzing User Preferences regarding Gamification Elements in Learning Management Systems." Proceedings of the 37th international conference on information systems (ICIS), Dublin, Ireland.
- Schöbel, S., M. Schmidt-Kraepelin, A. Janson, and A. Sunyaev. 2021. "Adaptive and Personalized Gamification Designs: Call for Action and Future Research." *AIS Transactions on Human-Computer Interaction* 13 (4): 479–494. <https://doi.org/10.17705/1thci.00158>.
- Sliwinski, J., M. Katsikitis, and C. Jones. 2017. "A Review of Interactive Technologies as Support Tools for the Cultivation of Mindfulness." *Mindfulness* 8 (5): 1150–1159. <https://doi.org/10.1007/s12671-017-0698-x>.
- Soh, H. L., R. C. Ho, C. S. Ho, and W. W. Tam. 2020. "Efficacy of Digital Cognitive Behavioural Therapy for Insomnia: A Meta-analysis of Randomised Controlled Trials." *Sleep Medicine* 75 (1): 315–325. <https://doi.org/10.1016/j.sleep.2020.08.020>.
- Tang, W. S., T. J. Ng, J. Z. Wong, and C. S. Ho. 2022. "The Role of Serious Video Games in the Treatment of Disordered Eating Behaviors: Systematic Review." *Journal of Medical Internet Research* 24 (8): e39527. <https://doi.org/10.2196/39527>.
- Thiebes, S., S. Lins, and D. Basten. 2014. Gamifying Information Systems—a synthesis of Gamification mechanics and Dynamics. ECIS.
- Thomas, P. C., K. Curtis, H. W. W. Potts, P. Bark, R. Perowne, T. Rookes, and S. Rowe. 2024. "Behavior Change Techniques Within Digital Interventions for the Treatment of Eating Disorders: Systematic Review and Meta-Analysis." *JMIR Mental Health* 11: e57577. <https://doi.org/10.2196/57577>.
- Tondello, G. F., R. R. Wehbe, L. Diamond, M. Busch, A. Marczewski, and L. E. Nacke. 2016. "The Gamification User Types Hexad Scale." Proceedings of the 2016 annual symposium on computer-human interaction in play.
- Van Houdt, L., M. Millemcamp, K. Verbert, and V. Vanden Abeele. 2020. "Disambiguating Preferences for Gamification Strategies to Motivate pro-environmental Behaviour." Proceedings of the annual symposium on computer-human interaction in play.
- Verzosa, N., S. Greaves, R. Ellison, A. Ellison, and M. Davis. 2018. "Eliciting Preferences for 'Gamified' Travel Surveys: A Best-Worst Approach." *Transportation Research Procedia* 32:211–223. <https://doi.org/10.1016/j.trpro.2018.10.039>.
- Yuan, T., and Y. Guo. 2021. "Gamification Design of Health Apps for the Elderly Based on the Kano Model and Conjoint Analysis Method." International conference on human-computer interaction.
- Zhang, M., S. Heng, G. Song, D. S. Fung, and H. E. Smith. 2019. "Co-designing a Mobile Gamified Attention Bias Modification Intervention for Substance use Disorders: Participatory Research Study." *JMIR MHealth and UHealth* 7 (10): e15871. <https://doi.org/10.2196/15871>.

Appendices

Appendix A: Selection of the Investigated Game Design Elements

To decide which GDEs were used in our survey, we synthesised (1) the findings from extant literature reviews regarding gamification in DMHIs (Brown et al. 2016; Cheng et al. 2019; Hoffmann, Christmann, and Bleser 2017) and (2) the chosen GDEs of previous BWS studies on users' GDE preferences in the health-related context (Berger et al. 2023; Berger and Jung 2024; Schmidt-Kraepelin et al. 2019). Table A1 summarises the selected GDEs with their sources and the descriptions provided to participants in our survey.

Table A1. The selected game design elements for the BWS-based surveys.

GDE	Description	Sources					
		1	2	3	4	5	6
Avatars	An avatar represents the user through a virtual character and can be designed by users regarding their own preferences (Hoffmann, Christmann, and Bleser 2017).	✓		✓	✓		✓
Badges	Badges belong to digital rewards and can be unlocked by fulfilling a specific goal, which is normally outside the scope of the core activities (Hamari 2017).	✓	✓	✓	✓	✓	✓
Goals	Goals are often implemented as challenges or quests to orientate users with a clear path of desired tasks and activities (Sardi, Idri, and Fernández-Alemán 2017).	✓	✓			✓	✓
Leaderboards	A leaderboard displays users' rank with regard to their peers based on specific criteria, such as performance (Sardi, Idri, and Fernández-Alemán 2017).		✓	✓		✓	✓
Levels	Levels provide information on the advancement of users regarding the overall performance (Hoffmann, Christmann, and Bleser 2017).	✓	✓	✓		✓	✓
Points	Points are credits accruing from users' participation and performance (Sardi, Idri, and Fernández-Alemán 2017).	✓	✓	✓	✓	✓	✓
Progress	Progress refers to a mechanic that shows the progression of users' achievements in a task or activity (Thiebes, Lins, and Basten 2014).	✓	✓		✓	✓	✓
Social interaction	Social interaction enables users to communicate with each other, including but not limited to sharing activities, commenting on others' achievements, and supporting each other (Hoffmann, Christmann, and Bleser 2017).		✓	✓	✓	✓	✓
Stories	Stories give users the possibility to experience a	✓	✓	✓		✓	✓

(Continued)

Table A1. Continued.

GDE	Description	Sources					
		1	2	3	4	5	6
	sequence of narrative events while performing an activity or task (Schmidt-Kraepelin et al. 2019).						
Virtual goods	Virtual goods are a form of rewards in gamification and usually implemented as intangible objects that can be earned (Schmidt-Kraepelin et al. 2019).	✓	✓	✓		✓	✓

1: Brown et al. (2016) 2: Cheng et al. (2019); 3: Hoffmann, Christmann, and Bleser (2017); 4: Berger et al. (2023); 5: Berger and Jung (2024); 6: Schmidt-Kraepelin et al. (2019)

We had to limit our study to the ten most relevant GDEs to ensure feasibility because the number of required choice sets depends on the number of incorporated objects. In addition, we scrutinised the implemented GDEs in some of the most popular DMHI applications on the market (e.g. *Head-space*, *Sleep Cycle*, *Wysa*, *Mood Tracker: Self-Care Habits*, *Recovery Record*). We did not encounter any additional elements that we had not considered before. Oriented toward extant research and practice findings, this selection aims to produce comparable results of users' preferences in different contexts.

Appendix B: scenario introductions and mockups presented in the surveys

At the beginning of each survey, we provided an introductory scenario describing the respective application context with illustrative mockups of GDEs. These materials are presented below.

Mindfulness practice

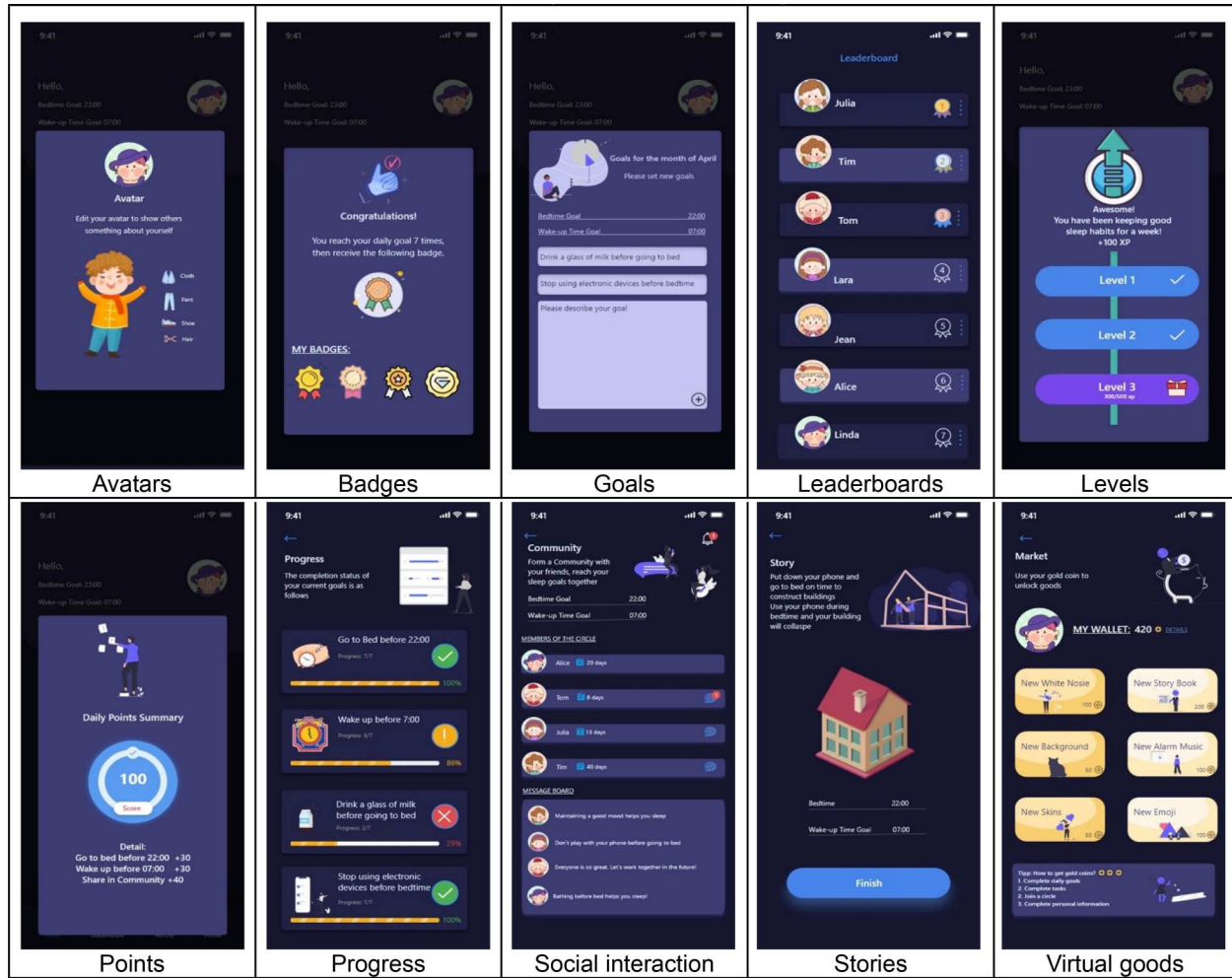
Imagine using a mobile app to learn and practice mindfulness and meditation. Within the app, you can follow various meditation classes that include building awareness, managing stress, or coping with adverse life events. Meditation classes are held by meditation experts. The app allows you to track your progress over time (e.g. amount of sessions, meditated minutes, daily mood) and supports its users in building a habit of meditating and being mindful. However, the app is not intended to accompany the therapy of specific mental health problems.



Sleep tracking

Imagine that you are using a mobile application to help you develop good sleep habits. You need to set a target time to go to sleep and wake up. At the target time you set for bedtime, the app will remind you that it's time to go to sleep, and each morning you need to open the app and complete a check-in when you wake up. The app will track your sleep

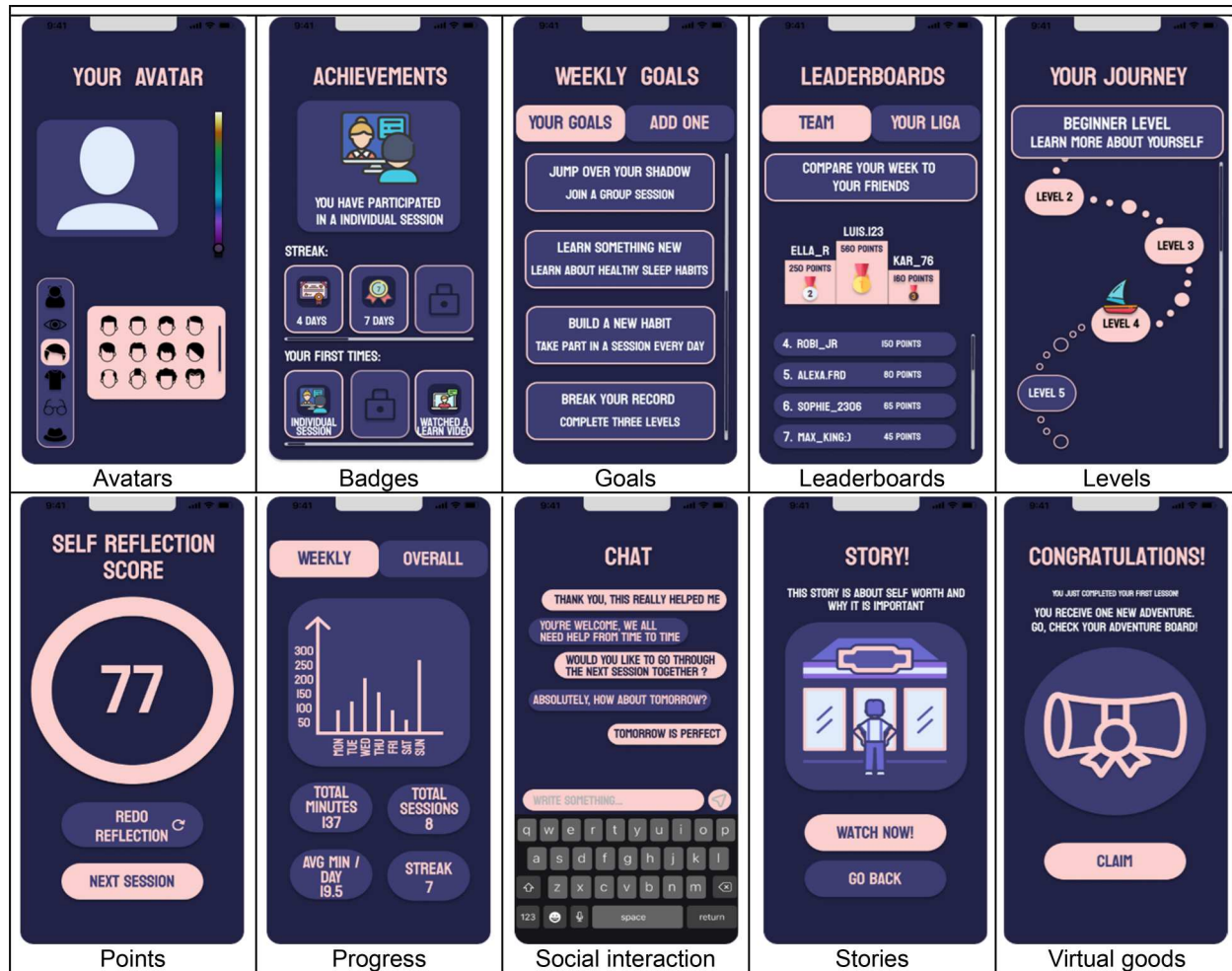
time and you can check your recent record at any time. It will also provide you with some aids (like white noise and bedtime stories) to help you fall asleep before you go to sleep. However, the app does not treat severe insomnia symptoms, but only helps you improve your sleep habits and reduce your dependence on your electronic devices before bedtime.



General mental health counselling

Imagine using a mobile app to address and improve your mental health. Within the app, you can have sessions with your therapist, participate in group sessions, or watch learning videos on

specific mental health issues. The app allows you to track your progress over time (e.g. number of sessions, type of sessions, learning effect after session). The different types of sessions are arranged online and then take place via a video call.



Mood tracking

Imagine using a mobile app to log your mood. In the app, you can write entries, get statistics on your mood, define personal goals, and archive your entries. The app allows you to record

your mood and track consistency over time. The app also allows you to identify triggers of bad mood or stress. Nevertheless, the app is not intended as a substitute for professional therapy for existing mental illnesses.



Eating disorder recovery

Imagine using a mobile app to manage your recovery from an eating disorder. Within the app, you can keep a record of meals, thoughts, and feelings, earn rewards, learn coping skills, and connect with other people on the same journey. The app allows

you to track your progress over time (e.g. daily logged meals, reached goals) and supports you in forming better eating habits and developing useful coping skills. The app is intended to accompany the therapy of eating disorders but doesn't claim to be a substitute for professional therapy.

