

RESEARCH ARTICLE OPEN ACCESS

A Pilot Study of a Gamified CBT-Based Digital Approach for Body Image Distress and Appearance-Related Concerns

Anushka Singh¹  | Naman Tyagi¹  | Fahad Eqbal¹  | Dolly Sharma¹  | Aikaterini Bourazeri² 

¹School of Engineering & Technology, Amity University Noida, Noida, India | ²School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

Correspondence: Aikaterini Bourazeri (a.bourazeri@essex.ac.uk)

Received: 1 April 2025 | **Revised:** 6 May 2026 | **Accepted:** 30 June 2026

Academic Editor: Nikhat Kaura

Keywords: body image distress | cognitive behavioral therapy | gamified intervention | mental health | serious games

ABSTRACT

Body image concerns related to perceived flaws in appearance represent a significant mental health challenge affecting a broad population. While Cognitive Behavioral Therapy (CBT) is an effective approach, stigma, financial costs, and limited access to trained professionals prevent many from seeking help. Digital interventions, particularly gamified approaches, offer an alternative by making support more engaging and accessible. To reflect the exploratory nature of this work, we developed a pilot-scale serious game that integrates CBT techniques with AI-driven analysis to provide interactive and personalized guidance. This pilot study provides preliminary indications of user-reported improvements in engagement and emotional reflection following gameplay. Compared to traditional approaches, the game's adaptive feedback system, informed by real-time emotional assessment, provides more personalized support during gameplay. The AI model used for analyzing user input achieved an accuracy of 97.15% on held-out test data, outperforming a BERT-based baseline classifier. While promising, these results are not intended to establish clinical efficacy; rather, they suggest that digital gamification may offer a scalable and accessible approach for supporting individuals experiencing body image distress. Future work will require larger samples, validated assessment tools, and multisite randomized controlled trials to evaluate effectiveness before potential clinical application, as well as exploration of augmented and virtual reality to further enhance user experience and supportive interaction outcomes.

1 | Introduction

Body image concerns related to perceived minor or invisible flaws in appearance are a prevalent and underdiagnosed mental health challenge, often imperceptible to others. Individuals experiencing such concerns may report chronic distress, compulsive behaviors such as mirror-checking or excessive grooming, and social withdrawal, all of which can significantly impair daily functioning. While the exact causes remain complex, research suggests that psychological, societal, and environmental factors interact to influence their onset and severity [1].

Modern societal dynamics including perfectionism, career pressures, and exposure to unrealistic beauty standards may

contribute to the development and exacerbation of appearance-related concerns [2, 3]. Social media, in particular, has been found to reinforce unattainable esthetic ideals, correlating with increased body dissatisfaction and heightened self-scrutiny [4–6]. However, beyond digital influences, broader lifestyle factors, including performance-driven cultures and the relentless pursuit of flawlessness, can also intensify psychological distress among affected individuals [7].

Despite growing awareness, appearance-related psychological distress remains underrecognized and undertreated. Cognitive Behavioral Therapy (CBT) remains an effective approach, demonstrating efficacy in restructuring negative thought patterns and reducing maladaptive behaviors [8]. However, stigma,

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Copyright © 2026 Anushka Singh et al. *Human Behavior and Emerging Technologies* published by John Wiley & Sons Ltd.

financial barriers, and the limited availability of trained professionals continue to hinder access to traditional therapy [9]. As a result, innovative digital approaches are increasingly explored as scalable alternatives.

Gamified mental health interventions have gained traction due to their engaging, interactive, and personalized approach to support delivery. Unlike traditional self-help applications, game-based approaches offer real-time feedback, motivation-driven engagement, and dynamic adaptation of support content, making them particularly suitable for addressing appearance-related concerns [10, 11]. This study introduces an AI-powered serious game that integrates CBT techniques with sentiment analysis to personalize supportive feedback based on user interactions. By detecting emotional variations in real time, the game provides adaptive CBT-informed prompts designed to encourage reflection on appearance-related thoughts and emotions [12]. This pilot study aims to conduct an initial feasibility evaluation of the proposed gamified approach and its potential to support engagement and reflection on body image concerns. Participants were not clinically diagnosed; instead, individuals were screened using a body image distress questionnaire to identify those experiencing relevant concerns.

Accordingly, this paper is structured as follows. Section 2 provides a comprehensive literature review, outlining existing research on digital mental health approaches and related interventions. Section 3 presents the theoretical foundation of this study, detailing the integration of CBT with game-based strategies. Section 4 describes the research design, including data collection, model development, and evaluation metrics. Section 5 presents the development of the serious game and its integration with AI-driven sentiment analysis for real-time adaptive support. Section 6 reports the empirical findings, examining the impact of the proposed approach on user engagement and emotional well-being. Section 7 outlines the study's limitations, including methodological, analytical, and clinical considerations. Finally, Section 8 concludes the paper with key insights and future research directions, including opportunities for enhancing the system through augmented and virtual reality.

2 | Background and Motivation

Understanding the scope of body-related concerns and their impact on individuals is essential for developing effective support strategies. Appearance-related distress represents a significant mental health challenge that affects a broad population, yet it often remains underrecognized due to various social and psychological barriers. Traditional approaches, while effective, often face challenges related to accessibility, adherence, and stigma. This section provides an overview of the prevalence and impact of body image concerns, followed by a discussion on existing support approaches and their limitations, highlighting the necessity for alternative, technology-driven solutions.

2.1 | Prevalence and Impact of Body Image Concerns

Body image-related distress affects a substantial proportion of the population and is associated with significant psychological

burden [13]. Despite its prevalence, such concerns often go unrecognized or unaddressed [14]. These experiences can significantly disrupt daily functioning and mental well-being [15], contributing to psychological distress, anxiety, and negative self-perception linked to perceived appearance flaws [16, 17]. Persistent focus on perceived imperfections may erode self-worth and contribute to maladaptive coping behaviors. This preoccupation can extend into work or social settings, impeding performance and engagement [18]. Individuals may engage in time-consuming behaviors such as excessive mirror-checking or grooming, further impacting daily life. The cumulative effect can result in reduced productivity, absenteeism, and broader psychosocial difficulties [19].

2.2 | Current Approaches and Limitations

CBT is widely recognized as an effective approach for addressing maladaptive thought patterns and behaviors related to body image concerns [9]. However, engagement with such approaches requires significant commitment, and stigma or reluctance to disclose appearance-related concerns may hinder help-seeking [10]. Pharmacological treatments such as selective serotonin reuptake inhibitors (SSRIs) may also be used in some cases [20] although responses vary across individuals [21]. Appearance-related distress often involves complex psychological factors, including low self-esteem and prior experiences, which may not be fully addressed by standard approaches [22]. These challenges highlight the need for more accessible, personalized, and scalable support strategies [23].

2.3 | Related Work

Growing research interest in appearance-related distress has led to extensive studies examining its prevalence, contributing factors, and impact. Veale et al. [24] conducted a systematic review of body image-related conditions, providing insight into their frequency across diverse settings and populations. Similarly, Alsaidan et al. [25] explored the prevalence of appearance-related concerns among young social media users, highlighting the influence of digital platforms on body image concerns and mental health outcomes. Schreurs and Vandenbosch [26] demonstrated how social media interactions contribute to self-objectification and body dissatisfaction among adolescents. Furthermore, McGrath et al. [27] provided a meta-analysis overview of prevalence patterns across populations.

Research has also highlighted challenges in identifying and addressing appearance-related distress. Krebs et al. [28] emphasized the need for improved recognition of body image-related difficulties, whereas Schneider et al. [29] explored subthreshold presentations and their psychological impact. These findings underscore the importance of early identification and supportive intervention strategies.

A range of approaches has been explored to support individuals experiencing body image-related distress. Singh and Veale [30] examined evidence-based psychological approaches, whereas Hong et al. [12] investigated alternative frameworks

such as acceptance and commitment-based strategies. Advances in digital technology have further enabled innovative approaches to improve accessibility and engagement. For example, Weingarden et al. [10] demonstrated the potential of smartphone-delivered CBT-based tools, whereas Summers et al. [11] utilized virtual reality to examine cognitive biases related to appearance perception.

These developments highlight the growing role of technology in supporting mental health. At the same time, ethical considerations such as privacy, data protection, and informed consent remain critical when deploying digital mental health tools. Cultural factors also play an important role in shaping body image perceptions and experiences, as demonstrated by Hakim et al. [31] and Santos et al. [32] emphasizing the need for inclusive and adaptable approaches.

Prior research provides a foundation for understanding appearance-related distress and the potential for technology-driven support strategies. This study addresses a critical gap by introducing a serious game designed to support reflection on body image concerns through interactive, CBT-informed principles. Unlike existing approaches, the proposed method incorporates sentiment analysis and predictive modeling to enhance personalization and engagement, offering a novel direction for digital mental health support.

3 | Theoretical Framework

This pilot study is grounded in the intersection of mental health, technology, and CBT, with a strong focus on addressing the complex challenges associated with body image distress. Recognizing the importance of early intervention, this research adopts a comprehensive, multidimensional approach, as illustrated in Figure 1 which outlines the framework of the proposed CBT-based serious game. To gather pertinent data for developing the game, an on-line questionnaire was administered to 50 voluntary participants, capturing crucial insights on demographics, quality of life, and body image concerns. These data provided a strong foundation for designing the game and customizing its gameplay components.

CBT principles provide a structured approach to understanding and addressing maladaptive thought patterns and behaviors related to body image concerns. These principles have been systematically incorporated into the proposed serious game:

1. Psychoeducation: providing essential information about body image concerns and their psychological effects to enhance awareness and early recognition.
2. Self-monitoring: encouraging individuals to track their thoughts, emotions, and behaviors related to body image concerns, aiding in identifying patterns and triggers.

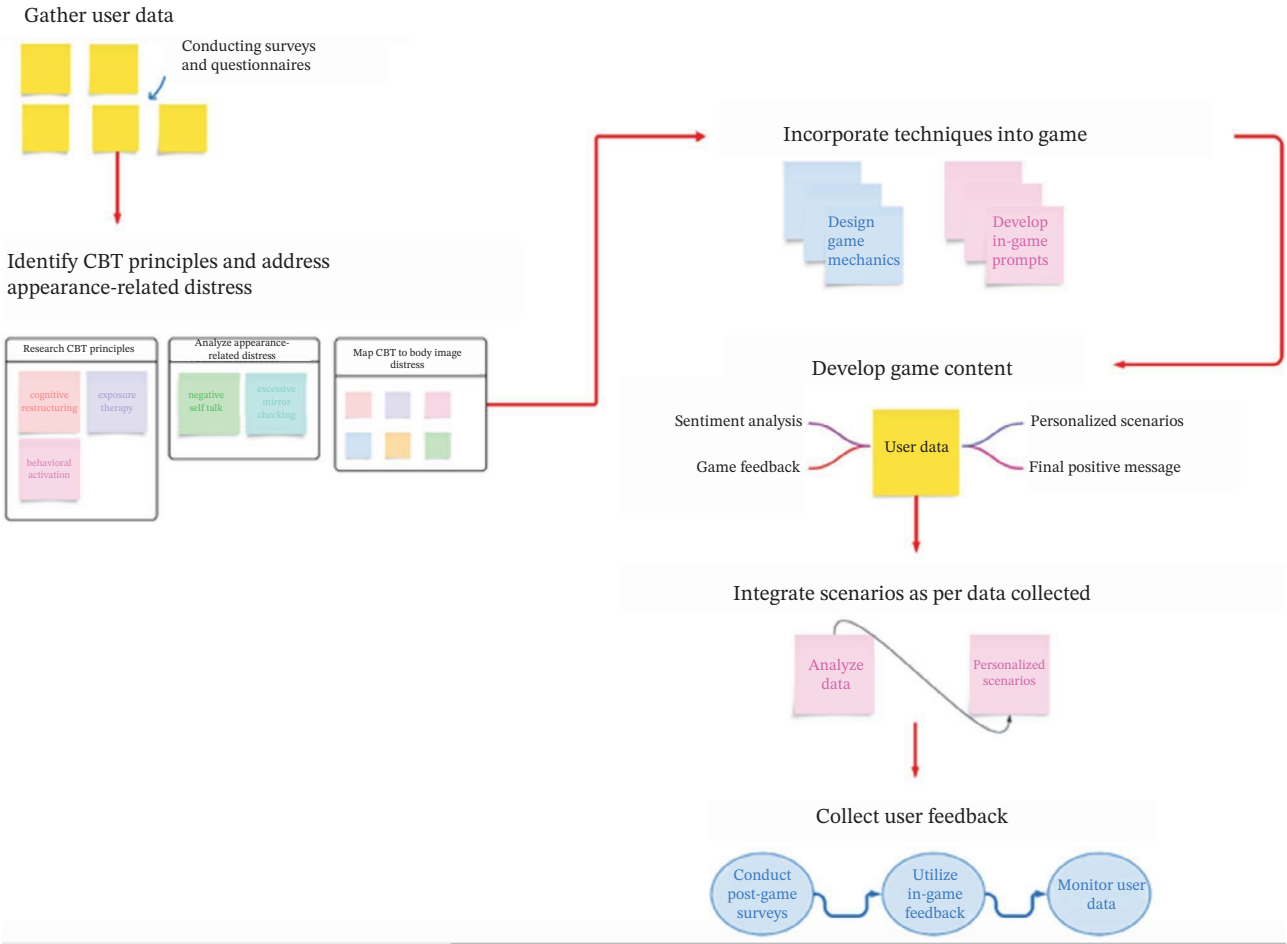


FIGURE 1 | Integration of CBT with gamification for body image support.

3. Cognitive restructuring: helping individuals recognize and challenge negative thought patterns and self-perceptions associated with their appearance.
4. Exposure and response prevention: gradual exposure to appearance-related fears while preventing compulsive behaviors that reinforce distress.
5. Behavioral experiments: encouraging participants to engage in real-world tests that challenge and reshape their distorted self-beliefs.

To further enhance engagement, the game employs effective gamification strategies, integrating interactive elements such as progress tracking, personalized feedback, and rewards based on user achievements. This ensures sustained motivation while reinforcing cognitive behavioral principles. The game is designed as an exploratory digital support tool, employing scoring mechanisms rooted in CBT principles to reflect patterns of distorted thought processes, engagement in repetitive behaviors, and levels of distress related to body image concerns. By analyzing user interactions, the game can dynamically adapt scenario difficulty and feedback based on the user's responses. Aligned with specific characteristics identified through questionnaire analysis, the game aims to offer a nuanced understanding of each participant's body image concerns, serving as both a self-reflection tool and a foundation for tailored gameplay strategies. This data-driven and CBT-informed approach facilitates alignment of game responses with individual user needs, supporting more personalized interaction. The primary goal of the proposed game is to support reflection on body image concerns and provide adaptive, CBT-informed feedback. This highlights the potential of integrating technology with psychological principles to develop innovative solutions for addressing appearance-related distress. By leveraging interactive gamification elements, users engage in structured CBT-based scenarios designed to encourage reflection on and challenge distorted thought patterns.

By employing a structured approach, this study helps bridge the gap between traditional mental health care and digital mental health solutions. The proposed serious game not only enables self-reflection but also fosters ongoing engagement, positioning it as a tool that may support early identification and reflection on body image concerns. The integration of gamification with established CBT principles signifies a substantial step forward in leveraging technology for mental health applications.

4 | Methodology

The methodology employed in this pilot study follows a mixed-methods approach, covering all stages from defining research objectives to developing a CBT-based serious game for body image distress: participant recruitment, data collection, preprocessing, model design, evaluation, and system integration.

4.1 | Participant Recruitment

Participants were recruited through online advertisements and outreach programs targeting individuals with potential body image concerns. All participants provided electronic

informed consent prior to participation after being presented with detailed information regarding the study's purpose, procedures, potential risks, and data handling practices. Participants were not required to have a formal clinical diagnosis; instead, they completed a prescreening questionnaire assessing levels of body image distress to ensure that individuals with relevant concerns were included in the study. This screening was not intended as a clinical diagnostic instrument, and future studies should incorporate validated assessment tools or clinician-administered diagnostic interviews. The final sample consisted of individuals aged between 18 and 45 years, with a gender distribution of 60% female, 35% male, and 5% nonbinary. Most participants were from English-speaking regions, allowing for a consistent linguistic dataset. A power analysis based on Cohen's d was conducted to ensure adequacy of the sample size. As the final sample exceeded the required threshold, the study was appropriately powered to detect medium-sized effects within the current exploratory pilot framework. The questionnaire dataset ($n = 50$) and the gameplay evaluation dataset ($n = 53$) were obtained from separate participant groups and correspond to different stages of the study (initial design and system evaluation).

4.2 | Ethics Approval and Participant Safety

This study was conducted in accordance with established ethical guidelines for research involving human participants, including the principles outlined in the Declaration of Helsinki. Ethical approval was obtained from the Ethics Sub-Committee 2 at the University of Essex under approval reference ETH 2324-1827. All participants provided electronic informed consent prior to participation, and participation was voluntary, with individuals free to withdraw at any time without penalty. Given that the study involved unsupervised interaction with CBT-informed, exposure-style digital content addressing sensitive body image concerns, a structured participant safety framework was implemented to mitigate potential psychological risk. This framework comprised preparticipation screening, explicit informed consent and risk disclosure, controlled and graded scenario design to minimize the likelihood of excessive distress, provision of mental health support resources, and poststudy debriefing.

Participants were screened using a body image distress questionnaire developed for this study, comprising self-report items assessing appearance-related concerns, distress levels, and related behaviors. This screening was used to identify individuals with relevant body image concerns and to flag elevated levels of distress. As the instrument was not a validated clinical diagnostic tool, it was not used to establish clinical diagnoses or apply formal clinical thresholds, but instead functioned as an initial safeguard within this exploratory study. Participants were explicitly informed in advance about the nature of the study, including the possibility of mild emotional discomfort when reflecting on body image concerns.

Gameplay scenarios were designed to follow a graded and low-intensity structure, avoiding highly distressing or triggering content. Participants were provided with information on how to access mental health support services and were advised to seek

professional help if distress increased during or after participation. All data were anonymized at the point of collection, and no personally identifiable information was stored alongside participant responses. Data were securely stored in controlled-access systems at the University of Essex and were accessible only to authorized members of the research team. Finally, participants were debriefed at the end of the study, including clarification of the exploratory nature of the system and its limitations as a nonclinical support tool.

4.3 | Data Collection

The dataset used in this study was derived from two primary sources: gameplay interaction data and mental health-related conversations. Gameplay data were collected from participants interacting with simulated CBT scenarios within the serious game, with all participants providing informed consent prior to data collection. During gameplay, user responses, sentiment scores, and emotional feedback were recorded, allowing the system to capture real-time interactions. Additionally, a curated dataset of mental health-related conversations between users and professionals was incorporated. This dataset, sourced from the “NLP Mental Health Conversations”¹ on Kaggle, contains 3512 anonymized conversational entries organized into two columns, Context and Response, representing user statements describing psychological concerns and corresponding supportive replies. This structure captures language patterns associated with emotional distress, self-perception, and coping behaviors commonly observed in online mental health discussions. Although the dataset is not specific to any single mental health condition, it includes expressions related to self-image and emotional distress, making it suitable for training models to detect sentiment patterns relevant to body image-related interactions. These examples were used to refine the game’s response generation, improving both its adaptability and the quality of interactions.

4.4 | Data Preprocessing

To process the textual data generated within the serious game, a structured preprocessing pipeline was implemented to enhance the accuracy of sentiment analysis and response generation. The text normalization process involved converting all text to lower-case, removing special characters, and eliminating unnecessary spaces to ensure consistency. Tokenization was then applied, segmenting text into smaller components to facilitate structured analysis. These tokens were transformed into numerical vector representations using pretrained word embeddings to enable the model to capture semantic relationships. Sentiment labeling was performed by assigning polarity scores to user responses based on predefined lexicons, classifying them as positive or negative. Finally, padding and truncation were employed to standardize input lengths, improving computational efficiency during model training. These steps ensured that the data were structured, noise-reduced, and optimized for sentiment classification.

4.5 | Model Architecture

The core of this study involves a custom-built machine-learning model designed to process gameplay interaction data and

generate responses aligned with CBT principles. The model architecture is structured to analyze user inputs, interpret sentiment, and adjust in-game responses accordingly. The model begins with dual input layers, processing user inputs from the game and the corresponding game-generated responses. These inputs are then transformed into numerical representations using shared embedding layers, enabling the system to recognize semantic relationships between words and phrases [33]. To analyze gameplay interactions effectively, temporal convolutional layers are applied to extract local features and patterns from user input sequences. Unlike traditional recurrent models, this approach allows the system to capture important context efficiently, making it suitable for real-time gaming environments [34].

A Dual Attention Mechanism prioritizes key contextual elements, ensuring that the most relevant aspects of the user’s input guide the response selection. The Cross-Stream Interaction Layer facilitates information exchange between user input processing and gameplay response generation, enabling dynamic adaptation to player interactions [35]. Inspired by transformer architectures [36], a Multi-Head Self-Attention component is incorporated to capture complex relationships within user-generated text, allowing the system to weigh the relative importance of different words effectively [37]. This supports that subtle variations in sentiment and phrasing are appropriately accounted for when generating responses. To ensure a consistent and personalized gameplay experience, a context memory component retains relevant details from prior interactions, allowing the system to adjust responses based on past user behaviors [38]. This mechanism enhances continuity in the game’s response strategy, making the gameplay more immersive and tailored to individual user needs.

The final output layer, activated by a sigmoid function, produces sentiment classifications that guide the game’s adaptive responses [39]. This architecture ensures that the system not only detects emotional shifts in player input but also generates adaptive responses informed by CBT principles. The overall architecture is illustrated in Figure 2 demonstrating how the different components interact to form an adaptive CBT-based support system.

4.6 | Sentiment Analysis Approach

Sentiment analysis is a central component of this methodology, serving as the foundation for the serious game’s adaptive feedback mechanism. By analyzing player interactions, the system dynamically adjusts in-game responses based on sentiment patterns, supporting a personalized and engaging user experience.

$$\text{Polarity} = (1/\text{Tot}(\omega)) \times [\sum(\omega \in \text{PosWords})(\text{SentimentScore}(\omega) \times \text{Freq}(\omega) \times \text{Intens}(\omega)) - \sum(\omega \in \text{NegWords})(\text{SentimentScore}(\omega) \times \text{Freq}(\omega) \times \text{Intens}(\omega))]$$

The polarity score is a measure of sentiment positivity or negativity within a given text input, aiding in interpreting emotional states throughout the game. The formula breakdown is as follows:

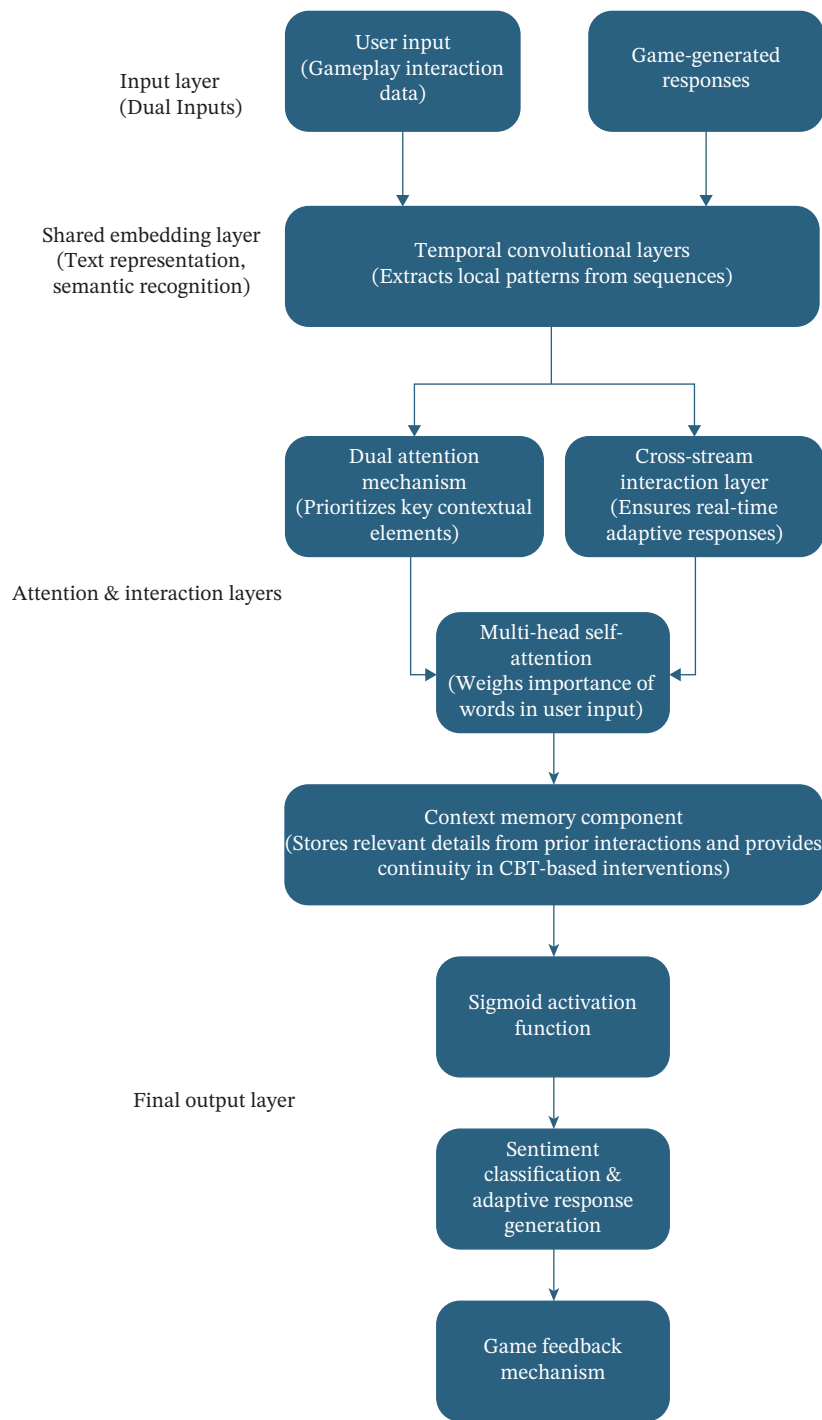


FIGURE 2 | Overview of the model architecture for CBT-based sentiment analysis in gameplay.

- $\sum \omega \in \text{PosWords}$: represents the summation over all positive words in the text.
 - $\text{SentimentScore}(\omega)$: refers to the predefined sentiment weight assigned to a particular word ω .
 - $\text{Freq}(\omega)$: accounts for how often a word appears in the text.
 - $\text{Intens}(\omega)$: measures the intensity of sentiment expressed by ω , differentiating between mild and strong emotional tones.
 - $\sum \omega \in \text{NegWords}$: is the summation over all negative words.
 - $\text{Tot}(\omega)$: represents the total number of words in the analyzed text.
- The sentiment analysis model operates using a lexicon-based approach to classify words as positive or negative based on predefined dictionaries. This process involves considering multiple parameters such as word frequency, sentiment intensity, and negative handling to improve classification accuracy. For

example, words such as “worst” or “horrible” carry a stronger negative weight than “bad.” Similarly, phrases containing negation, such as “not bad,” are correctly identified and adjusted to avoid misclassification as negative. Beyond individual word analysis, the system is capable of detecting qualitative shifts in sentiment by identifying patterns in user interactions. The presence of specific negative self-referential phrases such as “I am the ugliest person” allows the model to assess comparative self-evaluation tendencies. This helps in identifying potentially maladaptive self-talk, which can be flagged for targeted CBT-based support. Quantitative changes in sentiment polarity

scores indicate changes in emotional expression over time, allowing the game to adapt dynamically. Algorithm 1 summarizes the proposed lexicon-based sentiment analysis procedure used to preprocess user input, compute the polarity score, and assign the corresponding sentiment label. By leveraging this sentiment analysis framework, the serious game enhances real-time interaction capabilities, providing adaptive feedback tailored to each user’s evolving emotional state. This ensures that players receive personalized, context-aware responses aligned with CBT principles, making the gaming experience not only engaging but also supportive of emotional reflection.

ALGORITHM 1: Sentiment analysis with lexicon and polarity score.

Input: *user_text*, *Lexicon*, *Threshold_pos*, *Threshold_neg*

Output: *polarity*, *sentiment_label*

Preprocess(user_text) → *tokens*

Tot(ω) = $|\text{tokens}|$

pos_sum = 0

neg_sum = 0

for All tokens ω in *tokens* **do**

SentimentScore(ω) = *Lexicon*(ω)

Freq(ω) = *Frequency*(ω , *tokens*)

Intens(ω) = *IntensityModifier*(ω)

if *NegationContext*(ω) → True **then**

SentimentScore(ω) = *AdjustForNegation*(*SentimentScore*(ω))

end if

if *SentimentScore*(ω) > 0 **then**

pos_sum = *pos_sum* + *SentimentScore*(ω) * *Freq*(ω) * *Intens*(ω)

end if

if *SentimentScore*(ω) < 0 **then**

neg_sum = *neg_sum* + $|\text{SentimentScore}(\omega)|$ * *Freq*(ω) * *Intens*(ω)

end if

end for

polarity = (*pos_sum* − *neg_sum*) / *Tot*(ω)

if *polarity* ≥ *Threshold_pos* **then**

sentiment_label → Positive

else if *polarity* ≤ *Threshold_neg* **then**

sentiment_label → Negative

else

sentiment_label → Neutral

end if

return *polarity*, *sentiment_label*

Given the exploratory nature of this pilot study, a lexicon-based sentiment analysis approach was selected for its interpretability, computational efficiency, and suitability for early-stage prototyping.

4.7 | Model Training

To develop a sentiment classification model for the serious game, the dataset was divided into training and testing subsets using an 80–20 split. This allowed that the model was trained on a substantial portion of the data while retaining a test set for performance evaluation. Training was conducted using the Adam optimizer with a learning rate of 0.001, chosen for its dynamic adjustment of learning rates, which supports convergence. The binary cross-entropy loss function was employed to optimize classification between positive and negative sentiment labels. The model was trained in batches of 32 samples over 100 epochs, allowing it to generalize across different types of player interactions. To prevent overfitting, dropout layers were implemented, introducing randomness into the network and improving its generalization capability. Additionally, early stopping was employed, monitoring validation loss and halting training once no further improvement was observed. By incorporating these strategies, the training process facilitated the learning of meaningful sentiment patterns while maintaining adaptability for real-time gameplay interactions. The resulting model was designed to classify user sentiment, supporting the adaptive responses generated by the game.

4.8 | Evaluation Metrics

The model's performance was evaluated using a set of metrics to assess its performance in sentiment classification within the serious game. Accuracy measured the overall correctness of predictions, providing a high-level assessment of model performance. Precision and recall were used to evaluate the model's ability to classify positive and negative sentiments, supporting a balance between false positives and false negatives. The F1-score was used as a harmonized metric that integrates precision and recall, offering a more reliable evaluation in cases where class distribution is imbalanced. Additionally, the Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) score was used to assess the model's ability to distinguish between positive and negative sentiment. The ROC curve plots the tradeoff between the true positive rate (TPR) and the false positive rate (FPR) at different classification thresholds, whereas the AUC measures how well the model can separate classes. To further analyze classification performance, a confusion matrix was utilized to visualize misclassifications and identify areas for potential improvements. A correlation matrix was used for the questionnaire data to explore associations between psychological and behavioral attributes. To address the risk of Type I error arising from multiple correlations, p values for all pairwise correlations were evaluated using a False Discovery Rate (FDR) correction (Benjamini–Hochberg) across the full set of 45 correlations. Effect sizes were calculated for all inferential analyses to assess practical significance. Cohen's d was derived for independent t -tests, η^2 (eta-squared) for the one-way ANOVA, and Cramer's V for the chi-square test. Standardized regression

coefficients and R^2 were used to quantify effect magnitudes in the regression model.

4.9 | System Integration

The trained model was integrated into the backend of the serious game to enable real-time sentiment analysis and response generation. The system was developed to process player interactions, assess sentiment, and deliver adaptive responses informed by CBT principles. The frontend of the game featured an intuitive interface where participants engaged in interactive gameplay scenarios. User inputs were transmitted to the backend, where they underwent sentiment analysis to determine emotional tone and cognitive patterns. Based on the analysis, the game dynamically adjusted responses to provide personalized feedback. The sentiment classification component does not infer clinical states or deliver individualized therapy; rather, it supports the selection of predefined CBT-informed prompts within structured gameplay scenarios. Within this framework, sentiment classification outputs guided the selection of feedback prompts presented to the player. For example, responses indicating highly negative sentiment or self-critical language triggered reflective prompts encouraging players to examine potential cognitive distortions (e.g., magnification or selective attention to perceived flaws). When a player expressed a statement such as “Everyone will notice how bad my skin looks,” the sentiment classifier identified strong negative affect and self-critical language. The system then presented a prompt encouraging the player to evaluate evidence for and against this interpretation and consider alternative explanations, reflecting CBT-based cognitive restructuring techniques. In contrast, neutral or balanced responses resulted in supportive reinforcement or progression to the next scenario. A real-time feedback loop was implemented to log gameplay interactions, track sentiment changes over time, and refine response generation. This iterative learning process supported continuous improvement, allowing the system to adapt to evolving user behaviors and enhance system responsiveness.

5 | Game Development

5.1 | Serious Game Integration

The proposed serious game provides an engaging experience centered around body image distress, designed to enhance understanding of the complexities associated with appearance-related concerns. Through a series of diverse scenarios that reflect real-life challenges related to body image, players respond to prompts and make choices that foster reflection and understanding of appearance-related thoughts and behaviors. The narrative-driven design reinforces realistic interactions, supporting awareness and encouraging deeper reflection on body image concerns. Figure 3 provides two example in-game scenarios, illustrating how players interact with reflective prompts and receive supportive guidance within the game environment.

The game was developed using a structured set of CBT-informed scenarios reflecting key cognitive behavioral principles. The system includes three primary scenario types: (i) psychoeducation modules, introducing players to body image-related concerns,

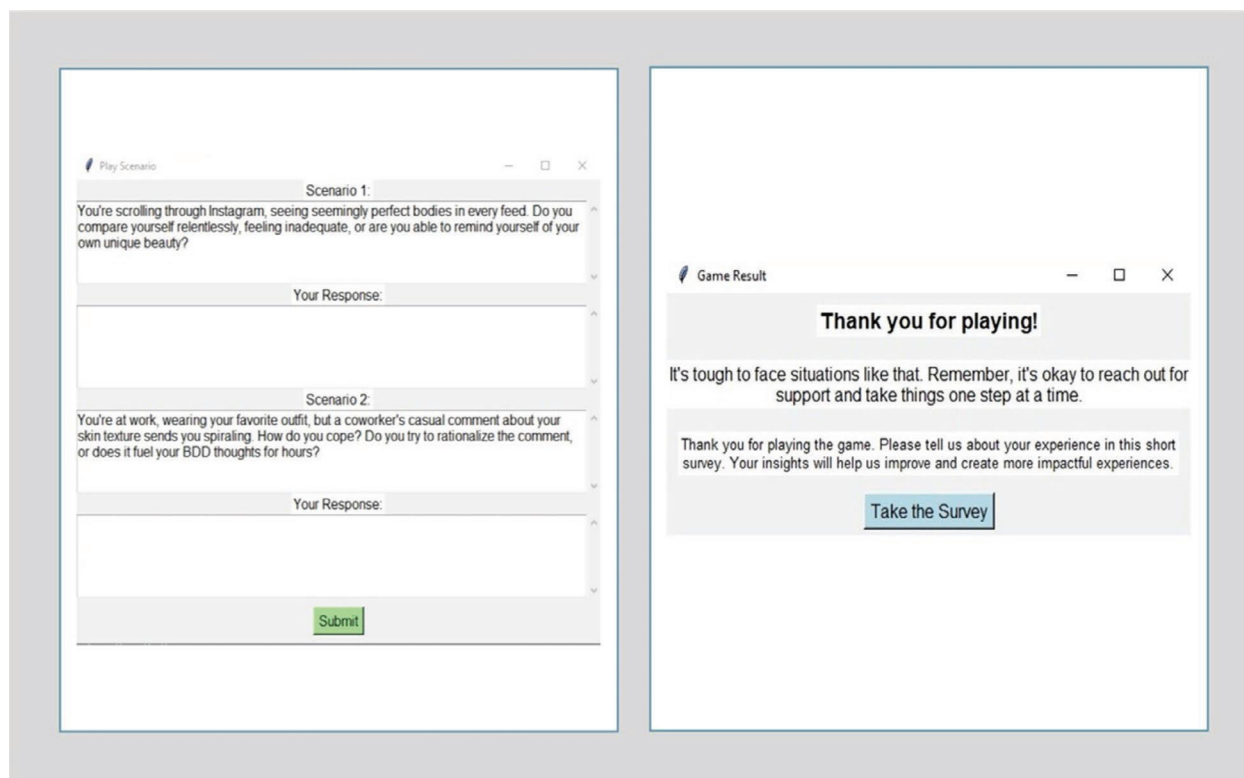


FIGURE 3 | An example of the game's user interface, illustrating how scenarios are presented (e.g., “You’re scrolling through Instagram, seeing seemingly perfect bodies in every feed. Do you compare yourself relentlessly, feeling inadequate, or remind yourself of your own unique beauty?”) and how players engage with reflective prompts. The figure highlights the structured presentation of scenario tasks and the supportive feedback style used throughout the system, showing how these interactions unfold during gameplay.

common cognitive distortions (e.g., magnification and selective attention to perceived flaws), and safety behaviors through short narratives and reflective prompts, (ii) cognitive restructuring tasks, where players respond to appearance-related automatic thoughts (e.g., “Everyone will notice how bad my skin looks today”), evaluate evidence for and against their interpretation, identify thinking patterns, and generate alternative, more balanced thoughts. Player choices influence the feedback delivered, reinforcing CBT principles, (iii) exposure and response-prevention (ERP) simulations, presenting graded challenges such as resisting mirror-checking, entering a social situation without camouflaging behaviors, or participating in a video call without adjusting appearance settings. These scenarios follow a graded-exposure hierarchy, and users' selected behaviors determine scenario progression.

For cognitive restructuring, players are presented with an automatic thought (e.g., “Everyone will stare at my skin during the meeting”), asked to identify the underlying distortion (such as catastrophizing), and then guided through evidence-for and evidence-against prompts before generating an alternative thought, which the system evaluates and reinforces. For ERP, scenarios encourage players to choose noncompulsive responses in situations that typically trigger appearance-related concerns. For example, when prompted with a mirror-checking urge, players can select between a habitual option (e.g., checking immediately) or an adaptive option (e.g., delaying or avoiding the check). Choosing the adaptive response leads to supportive CBT-informed feedback that reinforces tolerance of discomfort.

Psychoeducation modules introduce specific CBT principles through short narratives and reflective prompts, explaining selective attention to perceived flaws and asking players to apply it to a familiar daily situation.

Each gameplay session lasts approximately 10–15 min and includes three interactive scenarios designed to encourage reflection on appearance-related thoughts. Participants typically complete one full session, but the system supports repeated sessions with adaptive difficulty progression based on previous responses and sentiment trends. Distress was assessed immediately before and after a single gameplay session to capture short-term emotional change. Scenarios are constructed around everyday appearance-related challenges, including resisting grooming rituals, confronting perceived flaws, evaluating social interactions, and challenging distorted self-judgment. Each scenario uses a decision-based branching structure, allowing user choices to shape the game progression. The sentiment analysis model evaluates the emotional tone and informs the intensity and type of CBT-based feedback delivered (e.g., corrective guidance for maladaptive responses or reinforcement for adaptive choices). This structure supports an adaptive, personalized gameplay experience while maintaining alignment with established CBT principles.

In addition to scenario-based interactions, the game incorporates several gamification mechanisms that enhance engagement and support CBT principles. These include progress indicators that visually track the player's movement through

psychoeducation, cognitive restructuring, and exposure challenges; achievement badges awarded for completing tasks such as identifying cognitive distortions or engaging in response-prevention choices; and adaptive difficulty levels that progressively increase the complexity of cognitive tasks or the intensity of exposure scenarios. These mechanics support sustained motivation, reinforce adaptive behaviors, and facilitate repeated practice, which are key elements in CBT skill acquisition.

CBT principles were embedded directly into the game structure beyond sentiment-responsive feedback. Cognitive restructuring tasks require users to identify, evaluate, and reframe automatic thoughts before receiving AI-generated guidance. ERP tasks prompt users to engage in noncompulsive behaviors, reflect on emotional responses, and compare expected versus simulated outcomes, consistent with behavioral experiment principles. These structured cognitive behavioral components operate independently of sentiment analysis and support alignment with CBT methodology. To further clarify how CBT techniques are operationalized within the system, Figure 4 presents an illustrative backend example showing how user responses are analyzed and translated into CBT-informed feedback. The example demonstrates how the system processes inputs, assigns sentiment values, identifies cognitive patterns (such as distorted self-evaluations), and generates appropriate supportive feedback.

5.2 | CBT Integration Through Sentiment Analysis

This approach integrates CBT principles with sentiment analysis to support a more tailored and interactive user experience. The game begins with users engaging in carefully curated scenarios, where they provide responses that reflect their cognitive and emotional states. These inputs undergo sentiment analysis, classifying user statements as positive, negative, or neutral, and generating a sentiment score that reflects the emotional tone expressed. The sentiment scores and categorized outputs are securely stored in a database, along with user interaction data, enabling analysis of users' thought patterns and emotional responses over time. This data management system allows for ongoing tracking of interaction patterns, supporting an evolving understanding of user responses. The database interface displays sentiment categories, sentiment scores, and relevant interaction data, providing feedback aligned with users' emotional states.

This CBT-integrated, game-based approach examines the interplay between appearance-related concerns, repetitive checking behaviors, and emotional responses. By leveraging sentiment analysis and in-game behavioral tracking, the proposed system identifies patterns associated with emotional distress and self-evaluative thinking. Through real-time adaptation, the game adjusts feedback and scenario difficulty to introduce progressively challenging situations, encouraging users to reflect on and reconsider appearance-related thoughts. The incorporation of cognitive distortion analysis allows the game to respond to potentially maladaptive beliefs, reinforcing CBT-based coping strategies.

Additionally, by storing user interactions in a structured database, the system supports longitudinal analysis of

interaction patterns, enabling adaptive feedback over time. This integration supports ongoing user engagement and interaction, contributing to the development of digital support tools. Ultimately, this study contributes to the development of adaptive, personalized digital approaches for addressing body image concerns by combining game-based interaction, AI-driven sentiment analysis, and CBT principles. By enhancing the game's ability to capture and respond to user-specific cognitive and emotional patterns, this study supports the advancement of next-generation digital mental health support systems.

6 | Results

A series of statistical analyses was conducted to examine the patterns observed in the data. The analysis began with a power analysis using Cohen's *d* to assess whether the sample size was adequate for detecting effects of interest. Next, independent *t*-tests were used to examine sentiment trends in relation to text length. To further explore variations in sentiment, a one-way ANOVA was conducted to identify differences across various text features. Building on these findings, a multiple linear regression analysis was performed to explore potential predictors of sentiment classification. Finally, a chi-square test of independence assessed the association between text length categories and sentiment outcomes, providing additional insight into the relationship between textual characteristics and sentiment trends.

6.1 | Questionnaire Response Analysis

The demographics of the 50 participants ($n = 50$) who completed the initial questionnaire are summarized in Table 1, providing an overview of individuals reporting body image concerns. The findings indicated a diverse distribution, with females comprising 56% (28) of the sample, followed by males at 30% (15), and 14% (7) choosing not to disclose their gender. Participants below the age of 20 represented 12% (6), whereas 46% (23) fell within the 20–25 age bracket. Those aged between 26 and 31 accounted for 24% (12), and individuals above 31 comprised 18% (9) of the sample. Regarding marital status, approximately 66% (33) of participants identified as single, 8% (4) as married, whereas 26% (13) opted not to disclose their status. The higher proportion of participants in the 20–25 age group may provide contextual insight for the design of game scenarios, although this should be interpreted cautiously given the sample size. Figure 5 presents a correlation matrix illustrating the relationships among attributes extracted from the questionnaire. Analysis of the matrix revealed statistically significant associations between various dimensions of body-related concerns. To account for the number of correlations examined, a FDR correction (Benjamini–Hochberg) was applied across all 45 pairwise correlations. The strongest associations including the correlation between the impact on social life and preoccupation with perceived flaws ($r = 0.69$), appearance concerns and deliberate checking of specific features ($r = -0.49$), deliberate checking and related behaviors ($r = -0.42$), and related behaviors and first noticing appearance-related issues ($r = 0.40$) remained significant after correction. In contrast, weaker associations did not survive

```

"-Nu-IAAdMx9T48ZWYXZ2C": {
  "Name": " ",
  "Scenario1": "You're scrolling through Instagram, seeing seemingly perfect bodies in every feed. Do you compare yourself relentlessly, feeling inadequate, or are you able to remind yourself of your own unique beauty?",
  "Response1": "Yes I do compare them. Sometimes I end up being frustrated in myself and sometimes I end up being a little bit motivated to work on myself.",
  "Score1": -0.44375,
  "Scenario2": "You're having a good day, feeling confident, until a family member makes an insensitive comment about your appearance. Do you lash out in anger, or have you developed coping mechanisms to protect your emotional well-being?",
  "Response2": "Yes I would lash out in anger and tell them not to make any such comment about other people.",
  "Score2": -0.27499999999999997,
  "AverageScore": -0.359375,
  "FinalSentiment": "Negative",
},
"-Nu-6-NLmvtwdgAwlmW": {
  "Name": " ",
  "Scenario1": "You're at work, wearing your favorite outfit, but a coworker's casual comment about your skin texture sends you spiraling. How do you cope? Do you try to rationalize the comment, or does it fuel your BDD thoughts for hours?",
  "Response1": "I attempt to rationalize the comment, but it often lingers, fueling my BDD thoughts for hours. It's a struggle to overcome, especially when it targets something I'm already self-conscious about.",
  "Score1": 0.02,
  "Scenario2": "You're trying on clothes for an upcoming event, but nothing seems to hide your perceived flaws. Do you give up in frustration and cancel the event, or do you push yourself to find something that makes you feel good, even imperfectly?",
  "Response2": "Despite the frustration, I push myself to find something that makes me feel good, even imperfectly. Canceling the event would only reinforce negative feelings, whereas finding something that boosts my confidence, even if it's not perfect, empowers me to attend and enjoy the occasion.",
  "Score2": 0.06,
  "AverageScore": 0.04,
  "FinalSentiment": "Neutral",
},
"-Nu-ES8ZfryCR8c6Gzw0": {
  "Name": " ",
  "Scenario1": "You're trapped in a cycle of obsessive mirror checking, picking at perceived flaws, and engaging in unhealthy behaviors to fix yourself. Do you feel hopeless and alone, or have you found internal strength or external support to fight back against BDD's hold on your life?",
  "Response1": "Despite the challenging cycle, I've found internal strength to combat BDD's grip. With self-compassion and therapy, I'm gradually learning to appreciate myself beyond surface flaws and embracing my worth beyond physical appearance.",
  "Score1": 0.2,
  "Scenario2": "You're having a good day, feeling confident, until a family member makes an insensitive comment about your appearance. Do you lash out in anger, or have you developed coping mechanisms to protect your emotional well-being?",
  "Response2": "Having cultivated coping mechanisms, I choose to protect my emotional well-being. Instead of lashing out, I remind myself of my worth and focus on the positivity within me, understanding that others' comments don't define my self-worth.",
  "Score2": 0.15,
  "AverageScore": 0.175,
  "FinalSentiment": "Positive",
},

```

FIGURE 4 | Backend example illustrating how user inputs are processed and mapped to CBT-informed feedback. The system assigns sentiment scores, detects cognitive and emotional patterns, and selects appropriate supportive responses.

adjustment, indicating that only the most robust relationships persisted beyond the effects of multiple testing. Overall, the pattern of correlations suggests that cognitive, emotional, and behavioral aspects of body image distress tend to co-occur. Individuals reporting greater social interference tended to experience more frequent and intrusive thoughts about perceived

flaws. Appearance concerns were moderately associated with engagement in monitoring behaviors, suggesting that cognitive distress is linked to checking tendencies. Related appearance-focused behaviors also showed meaningful associations with both the initial noticing of perceived flaws and with checking patterns, reflecting interconnected behavioral responses.

TABLE 1 | Demographic characteristics of questionnaire participants (n = 50).

Variable	Category	n	%
Gender	Female	28	56%
	Male	15	30%
	Not disclosed	7	14%
Age	< 20	6	12%
	20–25	23	46%
	26–31	12	24%
	> 31	9	18%
Marital status	Single	33	66%
	Married	4	8%
	Not disclosed	13	26%

Figure 6 provides a comprehensive overview of the correlation matrix, emphasizing the consistency of the observed associations. The observed patterns highlight the relationship between appearance concerns and emotional responses related to perceived flaws. The findings also indicate an interplay between preoccupation with perceived flaws and observable behaviors, reinforcing the multifaceted nature of body image–related experiences. Additionally, the results suggest that individuals who avoid triggering situations may experience fewer negative emotions, indicating the presence of potential coping strategies in managing body image distress.

6.2 | Gameplay Data and Sentiment Analysis Evaluation

A separate group of 53 participants engaged with the serious game, which incorporated CBT principles and sentiment analysis techniques to examine emotional expressions and user engagement. Analysis of database records revealed that 75% (40) of the participants self-identified as female, whereas 25% (13) identified as male. This gender distribution indicates

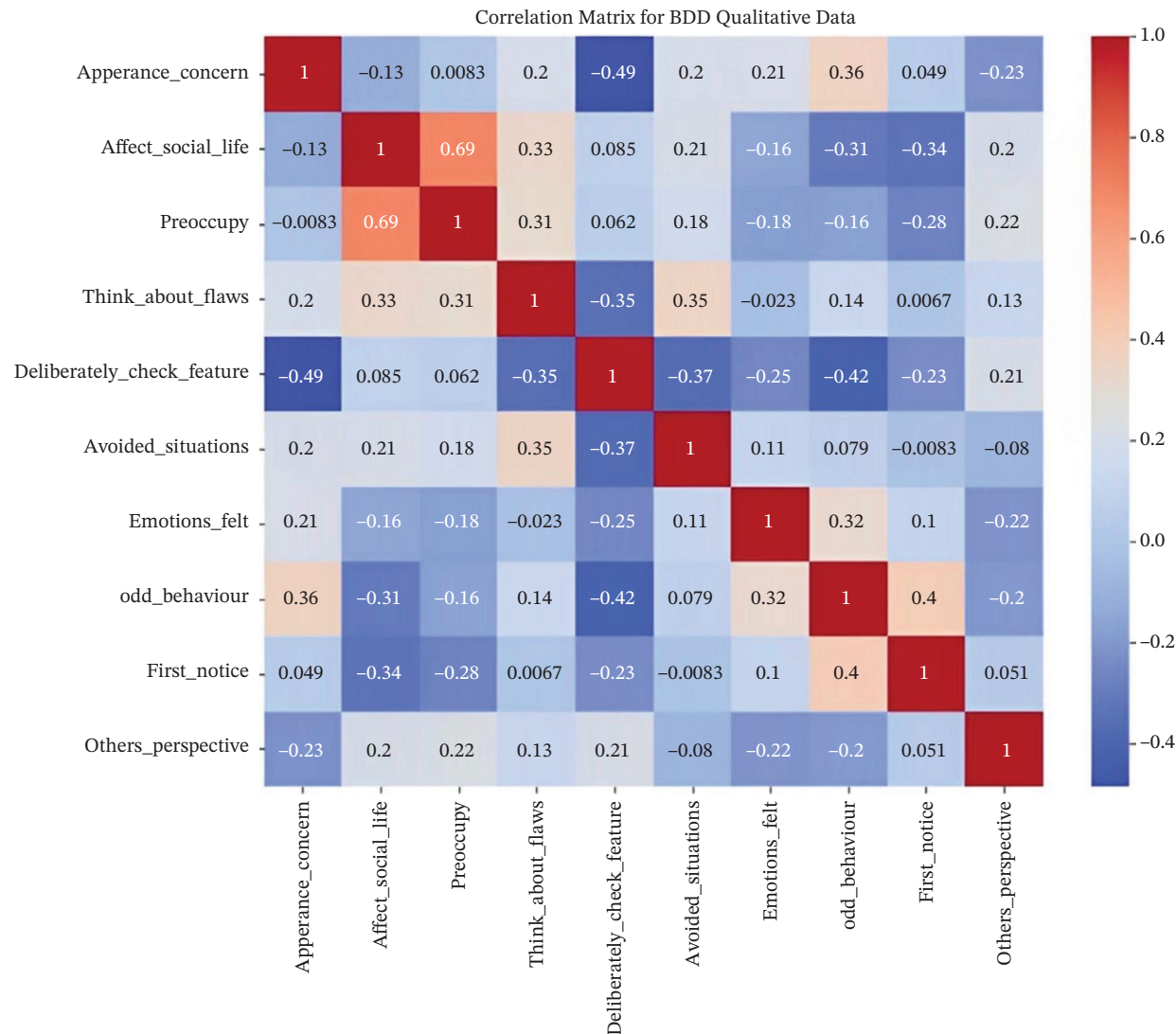


FIGURE 5 | Correlation matrix of psychological and behavioral attributes from BDD questionnaire.

Correlation	Description	r	p-value
Appearance Concern - Deliberately Checking Features	Positive association between concern and checking behaviors	0.75	< 0.001
Appearance Concern - Emotions Felt	Positive association between concern and negative emotions	0.68	< 0.001
Preoccupy - Think About Flaws	Positive association between preoccupation and flaw-focused thoughts	0.82	< 0.001
Preoccupy - First Notice	Negative association between preoccupation and remembering the onset of flaws	-0.51	< 0.05
Affect Social Life - Others' Perspective	Negative association between social impact and believing in others' perspective	-0.62	< 0.001
Avoided Situations - Emotions Felt	Negative association between avoidance and experiencing negative emotions	-0.59	< 0.01

FIGURE 6 | Correlation matrix with statistical significance (*r* and *p* values) across key measures.

a predominance of female participants within the age range of 19–35. A comprehensive evaluation of the user experience survey provided insights into user perceptions and reported emotional impact. The participant quotations presented below are included as illustrative examples of user feedback to provide contextual insight into user experiences. These quotations were not subjected to formal qualitative coding or thematic analysis. Participant feedback suggests generally positive responses to the game, with several highlighting its role in supporting reflection and self-expression. One participant remarked, “I feel good as I got a way to speak about my thoughts, and the scenarios given were quite intuitive and comprehensive,” illustrating how the game may provide a space for expressing thoughts and experiences. Similarly, another participant stated, “It feels nice to speak about your feelings,” indicating that the interaction may support open emotional expression. Additionally, comments such as “I feel confident” and “Felt positive about my body image” suggest that some participants perceived the interaction as supportive when engaging with body image-related concerns. Further analysis examined participants’ perceptions of the game experience, as illustrated in Figure 7. A substantial proportion of participants (77%) rated the game positively (score ≥ 4) in relation to supporting reflection on body image concerns. Moreover, 60% of participants reported a more positive perspective on body image immediately after gameplay. Additionally, 62% of participants expressed a likelihood of recommending the game to others, reflecting reported satisfaction and engagement.

Beyond qualitative feedback, sentiment analysis was applied to in-game textual responses to further examine patterns in emotional expression. One key aspect of this analysis involved

evaluating text length distribution across sentiment categories, as variations in text length may provide insight into differences in expression and engagement. Figure 8 illustrates the distribution of text lengths across sentiment labels (0 and 1), highlighting differences in response length based on sentiment classification. The presence of outliers suggests that certain responses, particularly in Category 1, exhibited greater text lengths.

Furthermore, an analysis of text length variability across sentiment classes was conducted to examine differences in response structure and distribution patterns. As shown in Figure 9, both sentiment classes (0 and 1) exhibited a wide range of text lengths, with several outliers extending beyond the main distribution. The violin plot illustrates that text length distributions are broadly similar across both sentiment classes, although some longer responses were more frequent in Sentiment Class 1. These findings are consistent with prior research suggesting that emotional intensity can influence response length, as stronger emotions (both positive and negative) may lead to more elaborated textual expressions [40]. The presence of outliers in both categories further indicates that some participants provided longer responses, regardless of sentiment classification.

Overall, these findings highlight relationships between sentiment expression, text complexity, and response patterns within the game environment. The integration of sentiment analysis with gameplay data offers preliminary insights into linguistic patterns of emotional expression and variability. This analysis demonstrates the potential of AI-driven approaches to support personalized interaction and reflective engagement for individuals experiencing body image concerns.

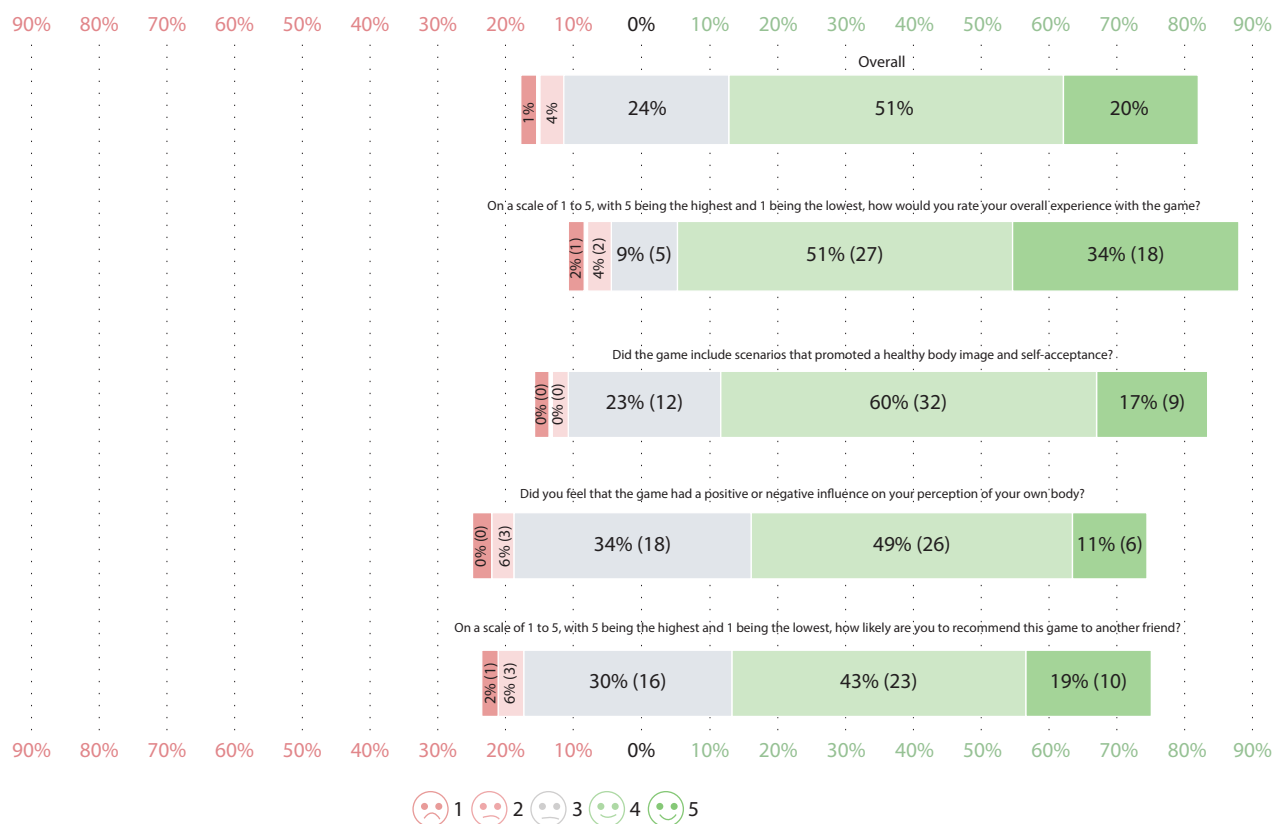


FIGURE 7 | User experience feedback from gameplay data (n = 53). Responses were collected using a 5-point Likert scale, and bars represent the percentage and count of responses for each rating category.

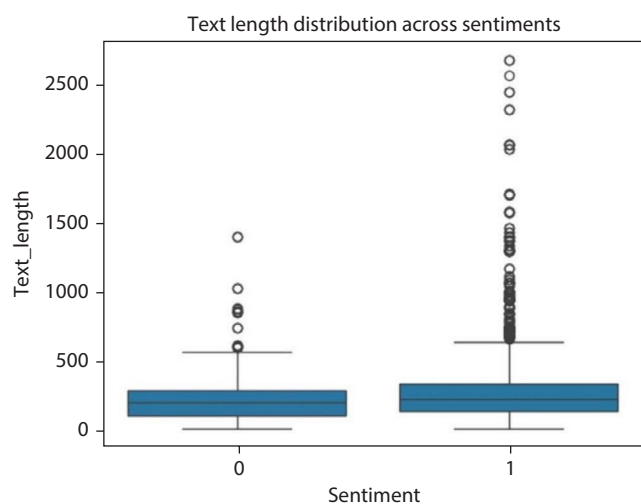


FIGURE 8 | Distribution of text lengths across sentiment categories.

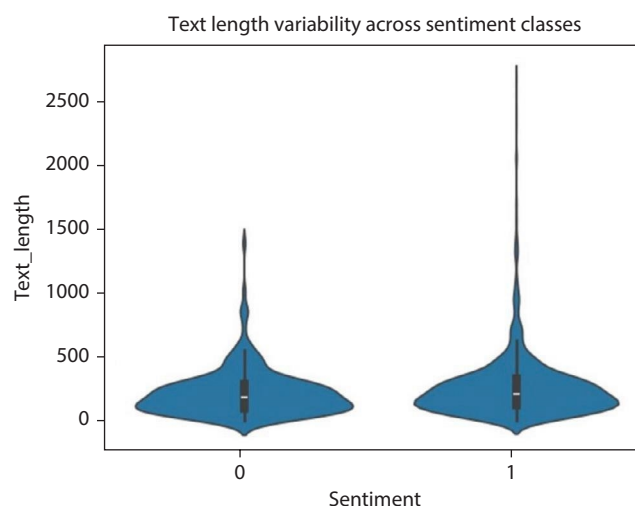


FIGURE 9 | Variability in text length across sentiment classes.

6.3 | Model Performance

The custom machine learning model demonstrated strong sentiment classification performance, achieving an overall accuracy of 97.15%. Throughout the training process, the model showed stable learning trends, with no clear indications of overfitting. By the final epoch, the model reached a training accuracy of 97.2%, whereas validation accuracy stabilized at 95.7%, closely aligning with the training results. The loss curves showed a consistent decline across both training and validation phases, indicating convergence and generalization

to unseen data. These performance metrics reflect classification accuracy within the study dataset but do not imply clinical validity or therapeutic effectiveness. Figure 10 illustrates these trends, depicting the training and validation accuracy over successive epochs.

To assess classification performance, key metrics were evaluated. The model achieved a precision score of 0.96, a recall of 0.98, and an F1-score of 0.97, indicating a balanced ability to classify sentiments while limiting false positives and false negatives. Its ability to distinguish between positive and

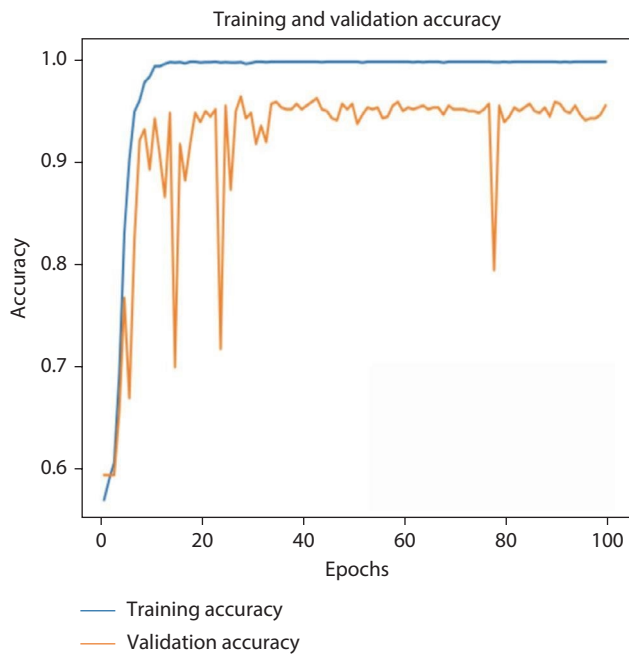


FIGURE 10 | Training and validation accuracy.

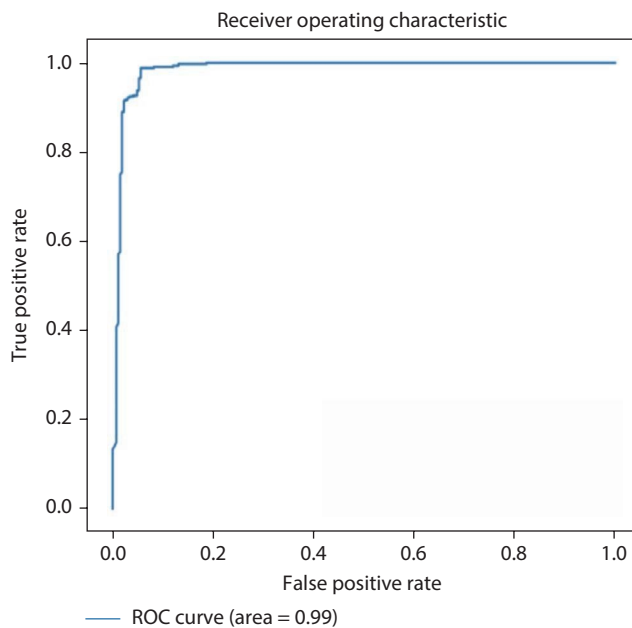


FIGURE 11 | ROC curve.

negative sentiment classes was further reflected by an ROC-AUC score of 0.98. The ROC curve, illustrated in Figure 11, highlights the trade-off between sensitivity and specificity across classification thresholds.

Additionally, benchmarking against pretrained sentiment analysis models, including a BERT-based sentiment classifier, indicated that the custom model achieved higher precision (0.96 vs. 0.89) and recall (0.98 vs. 0.85). Table 2 presents the performance comparison between the proposed model and the BERT-based baseline across the evaluated metrics. These findings suggest that the custom model performs well on this dataset, particularly in handling context-specific and nuanced sentiment expressions.

TABLE 2 | Performance comparison between the proposed sentiment classification model and a BERT-based baseline using precision, recall, and F1-score.

Model	Precision	Recall	F1-score
Proposed model	0.96	0.98	0.97
BERT-based classifier	0.89	0.85	0.87

While BERT-based classifiers perform strong on general sentiment tasks, they were less effective in this study, potentially due to the context-specific and emotionally nuanced nature of the language used. Participants frequently used short, fragmented, or highly personalized statements that pretrained models, optimized on large generic corpora, may misclassify. In contrast, the custom model was designed to incorporate task-specific linguistic features relevant to appearance-related distress and cognitive patterns, which may contribute to its performance on this dataset. Moreover, the lightweight architecture offers interpretability and computational efficiency, supporting integration into real-time interactive systems.

Overall, the empirical results and comparative analyses indicate that the model performs effectively for sentiment classification within this study. Its ability to maintain high classification accuracy while generalizing to unseen data suggests potential suitability for real-time sentiment analysis and adaptive feedback within the game environment.

6.4 | Power Analysis

A power analysis using Cohen's d was conducted to evaluate whether the sample size ($n = 50$) was adequate for detecting effects of interest, thereby reducing the risk of Type II errors. As an important component of experimental research, power analysis provides an indication of whether the study is sufficiently powered. A medium effect size ($d = 0.5$) was selected, following established benchmarks from previous research [41]. Given an alpha level of 0.05 and a statistical power of 80%, the analysis indicated that a minimum of 45 participants was required. As the study's sample size exceeded this threshold, it may be considered sufficient for detecting effects within the scope of this exploratory study.

6.5 | T-tests and ANOVA

Independent t -tests and one-way ANOVA were employed to investigate differences in key outcome measures across demographic groups, including age and gender. A one-way ANOVA was conducted to examine whether changes in reported distress scores varied across age groups. The results indicated a statistically significant difference across age groups ($F(2, 47) = 4.56$, $p = 0.015$, $\eta^2 = 0.16$). The effect size was in the medium range, suggesting that age was associated with variation in reported changes in distress scores. Post hoc pairwise comparisons using Tukey's HSD tests indicated that participants aged 20–25 reported greater changes in distress scores compared to participants aged 26–31 and those above 31 years ($p < 0.01$). These findings suggest age-related variation in responses; however,

they should be interpreted cautiously given the exploratory nature of the study and the limited sample size.

Independent *t*-tests were conducted to examine differences between males and females. Pregame scores did not show statistically significant differences between genders ($t(48) = 1.12$, $p = 0.26$), indicating comparable baseline levels across groups. The pregame comparison yielded a small effect ($r = 0.16$), consistent with the absence of meaningful baseline differences. Postgame scores indicated a statistically significant difference between females and males ($t(48) = 2.32$, $p = 0.02$), corresponding to a small-to-medium effect size ($r = 0.32$). These results indicate gender-related variation in reported scores; however, no causal conclusions can be drawn regarding differential engagement or outcomes. Observed differences may reflect variations in self-reporting patterns, engagement levels, or other unmeasured factors.

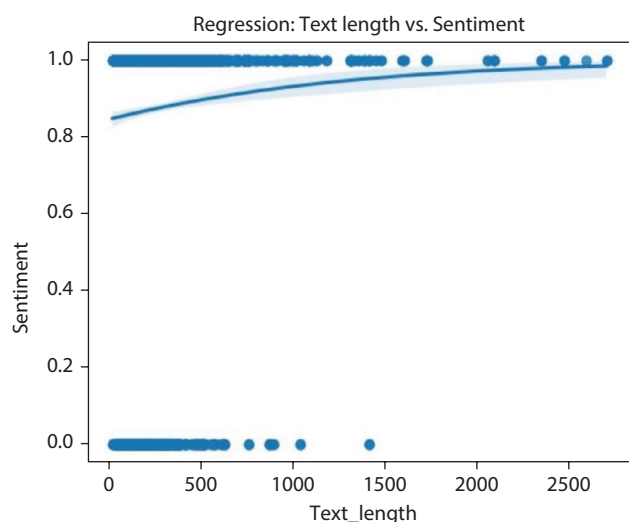


FIGURE 12 | Regression analysis of text length versus sentiment classification.

6.6 | Regression Models

To further explore factors associated with changes in reported distress scores, regression analyses were conducted to examine relationships between demographic variables (gender and age) and baseline distress levels. These analyses were intended to provide additional insight into how different factors may relate to observed changes in reported distress within this exploratory pilot dataset. The results indicated that higher baseline distress levels were associated with larger changes in reported distress scores. Age and gender also showed associations with variability in reported outcomes, although their effects were more modest, consistent with earlier findings suggesting that demographic factors may relate to differences in responses.

Additionally, to examine the relationship between text length and sentiment classification, a logistic regression model was applied. As shown in Figure 12, longer text responses were associated with a higher likelihood of being classified as positive sentiment (Label 1). Although the regression line suggests a weak increasing trend, the data exhibit high variability, indicating that while longer responses were more frequently associated with positive sentiment, text length alone is not a strong predictor of sentiment classification. These findings align with previous analyses of text length variability [41], further highlighting the complex relationship between emotional expression and textual features in sentiment analysis.

6.7 | Chi-Square Test

To investigate associations between categorical variables, chi-square tests were conducted to assess the relationship between text length category (short vs. long) and sentiment classification (0 vs. 1). The results, as shown in Figure 13, indicate a statistically significant relationship between text category and sentiment classification ($\chi^2(1, N = 50) = 5.67$, $p = 0.017$). The magnitude of this association was in the medium range

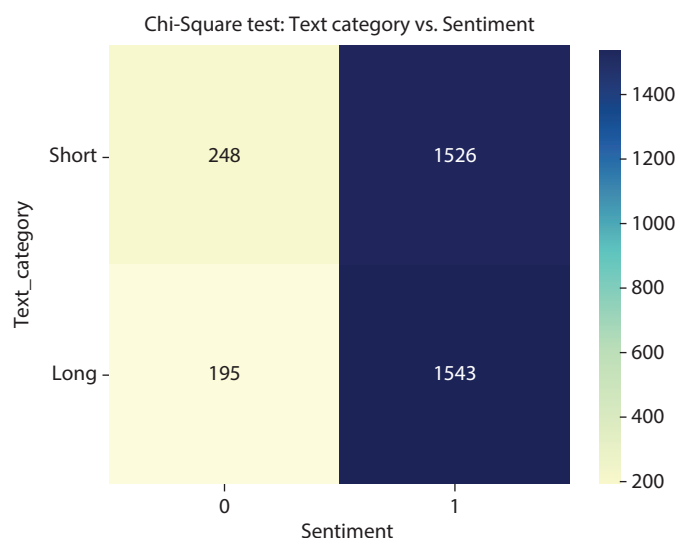


FIGURE 13 | Chi-square test analysis of text category (short vs. long) and sentiment classification.

(Cramer's $V = 0.34$), suggesting a moderate relationship between text length and sentiment classification. Specifically, a higher proportion of longer text responses were associated with positive sentiment ($n = 1543$), whereas shorter responses were more frequently classified as negative ($n = 248$). These findings suggest that text length may be associated with sentiment classification, aligning with previous analyses that indicated a weak positive trend between text length and sentiment in regression analysis [40]. However, while the chi-square results indicate a statistical association, they do not establish causation. The distribution of sentiment across text categories further highlights the complexity of sentiment expression, where multiple linguistic factors beyond length may contribute to classification outcomes.

6.8 | Summary of Key Findings

The statistical analyses indicated that text length is associated with sentiment classification, although effect sizes suggest a modest relationship. Independent t -tests and ANOVA identified differences in sentiment across text length groups, whereas regression analysis showed that text length alone explains only a small proportion of the variance in sentiment scores. A statistically significant chi-square test further indicated an association between sentiment classification and text length categories, highlighting the importance of considering multiple features in sentiment analysis.

Beyond sentiment classification, the statistical analyses, including regression models and chi-square tests, suggest that the system may support short-term emotional reflection related to body image concerns. The key findings include the following:

- Self-reported distress scores changed following gameplay across participants, with variations observed across age and gender groups.
- Baseline distress levels were associated with the magnitude of changes in reported distress scores, with larger changes observed among participants with higher initial distress.
- Demographic differences in reported outcomes suggest the potential value of personalized approaches, highlighting that responses may vary across age and gender groups.

Additionally, the AI-driven sentiment classification model demonstrated high performance on this dataset (accuracy: 97.15%, ROC-AUC: 0.98). These findings indicate the potential of the proposed approach as an exploratory AI-enhanced system for supporting interactive CBT-informed digital experiences.

7 | Limitations

Despite promising findings, this pilot study has several limitations that should be considered when interpreting the findings, including the absence of a control group, reliance on short-term

pre-post self-report measures, the use of nonvalidated distress indicators, and limitations of lexicon-based sentiment analysis. First, the sample sizes for both the questionnaire analysis ($n = 50$) and gameplay evaluation ($n = 53$) are modest and limit the generalizability of the results. The recruitment strategy relied on online advertisements targeting individuals who self-identified as having body image concerns, which may have introduced self-selection bias. Moreover, participants were not required to have a formal clinical diagnosis; therefore, the sample reflects individuals with self-reported body image distress rather than clinically verified conditions. This limits the clinical applicability of the findings and restricts their generalizability to clinically diagnosed populations.

In addition, the sample was predominantly young and female, a demographic pattern that emerged from the voluntary recruitment process. Although unintentional, this further limits generalizability, as body image concerns affect individuals across genders and age groups. As a result, the findings should be interpreted within the context of a community-based sample with varying levels of body image concern rather than a clinically verified and demographically representative cohort. Future research will incorporate recruitment through clinical services, validated diagnostic assessments, and stratified sampling methods to ensure broader demographic and clinical representation. Furthermore, the study did not include culturally tailored content or analyses. Although recruitment occurred internationally, participation was self-selected and resulted in a predominantly Indian sample, and the game was not explicitly adapted for cultural differences in the expression or experience of body image-related concerns. Consequently, the generalizability of findings across cultural contexts is limited. Future research will incorporate culturally informed design and cross-cultural validation to ensure broader applicability of the system.

Second, the absence of a control or comparison group limits the ability to draw causal conclusions about the game's effects. Changes observed following gameplay may also reflect natural variation, expectancy or placebo effects, or the general emotional impact of engaging with an interactive digital activity. Third, outcomes were derived from self-report measures collected immediately before and after a single interaction with the system, and no follow-up assessments were conducted. Consequently, it is not possible to determine whether the observed changes in distress were sustained over time or whether users would continue to engage with the game over longer periods. Longitudinal assessment is essential for evaluating the durability of change and understanding how digital CBT-based tools support ongoing emotional and behavioral patterns. Future research will therefore incorporate multi-timepoint follow-up to examine whether observed changes in self-reported distress are maintained over time.

An additional safety consideration pertains to individuals experiencing severe distress or crisis-level symptoms. As a pilot-stage digital tool, the game is not designed to replace professional diagnosis or treatment, nor is it suitable for unsupervised use by individuals with acute symptomatology, suicidality, or complex comorbid conditions. Because the current prototype lacks real-time clinical monitoring or risk-detection mechanisms, there is a possibility that high-risk users could engage with the system

without receiving appropriate professional support. Future development will therefore incorporate safeguards such as brief severity screeners prior to gameplay, automated prompts directing users to clinical services when high-risk responses are detected, and optional clinician-linked monitoring modules to ensure safe use.

Moreover, the study did not employ validated psychiatric rating scales (such as standardized measures of anxiety, depression, or functional impairment). As a result, the observed changes in distress should not be interpreted as clinical efficacy but rather as preliminary responses to a gamified digital interaction. The CBT-informed scenarios were developed based on established literature but were not formally reviewed by licensed clinical psychologists or domain experts, which may limit their clinical fidelity. Although the sentiment analysis model demonstrated high performance, the study does not assess whether this classification accuracy translates into improved CBT-related outcomes or user benefit. The model's role is limited to supporting feedback selection within predefined scenarios, and its contribution to therapeutic processes requires further evaluation in controlled and clinically grounded studies. Incorporating validated clinical instruments in future studies will be essential for establishing clinically meaningful change and enabling comparison with existing digital and therapist-delivered CBT approaches. Additionally, several statistical tests were conducted as part of this exploratory analysis, and while appropriate caution was applied in interpretation, the risk of Type I error remains. Accordingly, the analysis prioritized effect sizes and overall patterns rather than isolated *p* values. Future adequately powered studies will incorporate formal correction procedures to control for multiple comparisons.

Furthermore, the regression model accounted for 32% of the variance in reported distress changes, indicating that additional predictive factors were not captured in this early-stage analysis. This level of explained variance is typical in exploratory behavioral research, particularly when working with heterogeneous, nonclinical samples and a limited set of predictors. Unmeasured influences such as prior experiences, comorbid symptoms, coping styles, or environmental stressors likely contributed to individual differences in outcomes. Finally, the lexicon-based sentiment analysis approach may overlook nuanced or context-dependent emotional expressions, which could lead to misclassification of complex affective states. Although the sentiment analysis model achieved high accuracy within the study dataset, its performance may vary across broader linguistic, cultural, and real-world contexts. These limitations highlight the need for larger samples, multisite randomized controlled trials, and longitudinal designs in future research.

8 | Conclusions and Future Work

This pilot study conducted an initial evaluation of a CBT-based AI system delivered through a serious game, designed to support individuals experiencing body image concerns. Through scenario-based gameplay and sentiment analysis, the game provided an interactive and adaptive digital environment that supported self-reflection and emotional expression. The statistical analyses, including regression models, chi-square tests, and ANOVA, indicated changes in self-reported distress scores

immediately following gameplay, suggesting potential benefits while not establishing clinical effectiveness. Variations in reported outcomes were observed across demographic groups, including age and gender.

Additionally, baseline distress levels were associated with the magnitude of changes in reported distress scores, suggesting that individual differences may play a role in how users engage with the system. These findings highlight the potential value of considering user variability when designing digital mental health tools. The AI-driven sentiment classification model demonstrated high performance on the study dataset (accuracy; 97.15%, ROC-AUC: 0.98), indicating its suitability for sentiment analysis within this context. In addition to the sentiment polarity and text-length features examined in this pilot study, future research will expand the analysis of linguistic and behavioral markers to capture deeper cognitive and emotional processes. Potential features include lexical diversity, frequency of cognitive processing terms (e.g., causal, insight-oriented, or self-referential language), emotional granularity, and patterns of in-game behavioral engagement. With larger and more diverse datasets, exploratory factor analysis may be conducted to identify latent dimensions underlying user responses, and qualitative methods may help contextualize themes related to appearance concerns and coping strategies.

The proposed game may be conceptualized as a complementary digital tool that could support existing mental health practices rather than replacing professional services. Within a stepped-care framework, the system may serve as a low-intensity engagement and psychoeducation tool, supporting early exploration of body image concerns and facilitating reflection. These potential applications require further validation through controlled and clinically informed studies.

Building upon these findings, future work will focus on expanding the technological capabilities and accessibility of the game. The integration of Augmented Reality (AR) and Virtual Reality (VR) presents opportunities to enhance user engagement and create more immersive interactive experiences. AR may enable users to interact with sentiment-based features in a private and flexible manner, whereas VR may support simulated social scenarios for experiential learning. These developments introduce important technical and ethical considerations, including data privacy, informed consent in immersive environments, accessibility, and sustained user engagement. Future development will incorporate safeguards to ensure that these capabilities remain ethically appropriate.

Additional directions include the use of adaptive AI methods such as reinforcement learning and longitudinal sentiment modeling to adjust system responses based on user interaction patterns. Context-aware models (e.g., transformer-based architectures or hybrid lexicon-machine learning systems) may also be explored to improve detection of nuanced emotional expressions. These developments may help address limitations of the current lexicon-based approach. Future studies will incorporate longitudinal designs, larger and more diverse samples, and randomized controlled trials to evaluate effectiveness and generalizability. Expanding demographic diversity will further clarify

how different populations interact with the system and inform necessary adaptations.

In conclusion, this pilot study highlights the potential of AI-enhanced CBT-informed digital approaches to support individuals experiencing body image concerns. The integration of sentiment analysis, gamified environments, and adaptive feedback mechanisms demonstrates a promising direction for exploratory digital mental health support. Further research is required to evaluate effectiveness, validate outcomes using standardized measures, and determine applicability in broader and clinically defined populations.

Author Contributions

Anushka Singh: software, investigation, data curation, writing – original draft. Naman Tyagi: software, investigation, data curation, writing – original draft. Fahad Egbal: software, investigation, data curation, writing – original draft. Dolly Sharma: conceptualization, methodology, supervision, writing – review and editing. Aikaterini Bourazeri: conceptualization, methodology, supervision, formal analysis, project administration, writing – original draft, writing – review and editing.

Funding

No funding was received for this manuscript.

Disclosure

All authors reviewed the manuscript, contributed to the interpretation of the results, and approved the final version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are not publicly available due to privacy considerations but are available from the corresponding author upon reasonable request. The data are stored in a controlled-access repository at the University of Essex. To support reproducibility, the nonsensitive components of the study are publicly available in the following GitHub repository:

<https://github.com/anushkasingh03/BDD>.

Endnotes

¹NLP Mental Health Conversations Dataset

References

1. J. D. Feusner, F. Neziroglu, S. Wilhelm, L. Mancusi, and C. Bohon, "What Causes BDD: Research Findings and a Proposed Model," *Psychiatric Annals* 40, no. 7 (2010): 349–355, <https://doi.org/10.3928/00485713-20100701-08>.
2. M. Vismara, V. Caricasole, V. Starcevic, et al., "Is Cyberchondria a New Transdiagnostic Digital Compulsive Syndrome? A Systematic Review of the Evidence," *Comprehensive Psychiatry* 99 (2020): 152167, <https://doi.org/10.1016/j.comppsy.2020.152167>.
3. P. T. S. Sulistyo, M. E. Sukanto, and N. Ibrahim, "Social Media Pressure and the Body Dysmorphic Disorder Tendency in Women: The Mediating Role of Perfectionism," *Psikohumaniora: Jurnal Penelitian*

Psikologi 7, no. 2 (2022): 137–152, <https://doi.org/10.21580/pjpp.v7i2.10637>.

4. H. K. Jarman, M. Fuller-Tyszkiewicz, S. A. McLean, et al., "Who's Most At Risk of Poor Body Image? Identifying Subgroups of Adolescent Social Media Users Over the Course of a Year," *Computers in Human Behavior* 147 (2023): 107823, <https://doi.org/10.1016/j.chb.2023.107823>.
5. F. Gioia, M. D. Griffiths, and V. Boursier, "Adolescents' Body Shame and Social Networking Sites: The Mediating Effect of Body Image Control in Photos," *Sex Roles* 83, no. 11–12 (2020): 773–785, <https://doi.org/10.1007/s11199-020-01142-0>.
6. S. C. Tremblay, S. Essafi Tremblay, and P. Poirier, "From Filters to Fillers: An Active Inference Approach to Body Image Distortion in the Selfie Era," *AI & Society* 36, no. 1 (2021): 33–48, <https://doi.org/10.1007/s00146-020-01015-w>.
7. B. MacKenzie, A. Galpin, and P. White, *Unplugged: Evolve From Technology to Upgrade Your Fitness, Performance, & Consciousness* (Simon and Schuster, 2017).
8. E. E. Bernstein, A. Szkutak, A. Fang, A. J. Blashill, J. Ragan, and J. L. Greenberg, "Cognitive Behavioral Therapy for Body Dysmorphic Disorder," in *The Massachusetts General Hospital Handbook of Cognitive Behavioral Therapy* (Springer, 2023), 201–215, https://doi.org/10.1007/978-3-031-29368-9_14.
9. S. Khemlani-Patel, and F. Neziroglu, *Body Dysmorphic Disorder* (Hogrefe Publishing GmbH, 2022), <https://doi.org/10.1027/00500-000>.
10. H. Weingarden, A. Matic, R. G. Calleja, J. L. Greenberg, O. Harrison, and S. Wilhelm, "Optimizing Smartphone-Delivered Cognitive Behavioral Therapy for Body Dysmorphic Disorder Using Passive Smartphone Data: Initial Insights From an Open Pilot Trial," *JMIR mHealth and uHealth* 8, no. 6 (2020): e16350, <https://doi.org/10.2196/16350>.
11. B. J. Summers, A. C. Schwartzberg, and S. Wilhelm, "A Virtual Reality Study of Cognitive Biases in Body Dysmorphic Disorder," *Journal of Abnormal Psychology* 130, no. 1 (2021): 26–33, <https://doi.org/10.1037/abn0000563>.
12. K. Hong, V. Nezhgovorova, and E. Hollander, "New Perspectives in the Treatment of Body Dysmorphic Disorder," *F1000Research* 7 (2018): 361, <https://doi.org/10.12688/f1000research.13700.1>.
13. Y. Wang, Y. Chen, C. Lu, et al., "The Psychological Factors Mediating/Moderating the Association Between Body-Image Disturbance and Depression: A Systematic Review," *PsyCh Journal* 13, no. 4 (2024): 527–540, <https://doi.org/10.1002/pchj.754>.
14. G. K. Jonathan, C. Armstrong, P. Miyares, J. Williams, and S. Wilhelm, "Advancing Psychosocial Treatment for Body Dysmorphic Disorder: A State-of-the-Science Review," *Behavior Therapy* 55, no. 6 (2024): 1249–1288, <https://doi.org/10.1016/j.beth.2024.04.002>.
15. E. C. Wolfe, I. Snorrason, Z. E. Laky, et al., "Clinical Characteristics Among Sexual Minority and Heterosexual Women With Body Dysmorphic Disorder," *Body Image* 49 (2024): 101687, <https://doi.org/10.1016/j.bodyim.2024.101687>.
16. S. Sauer-Zavala, and D. H. Barlow, *Neuroticism: A New Framework for Emotional Disorders and Their Treatment* (Guilford Publications, 2021).
17. A. Berjaoui, and B. Chahine, "Body Dysmorphic Disorder Among Lebanese Females: A Cross-Sectional Study," *Journal of Cosmetic Dermatology* 23, no. 2 (2024): 591–599, <https://doi.org/10.1111/jocd.16003>.
18. C. Lian, and X.-X. Chen, "Does Beautification Technology Use Affect Appearance Anxiety? An Exploration of Latent Mechanisms," *Computers in Human Behavior* 146 (2023): 107793, <https://doi.org/10.1016/j.chb.2023.107793>.
19. S. E. Schwartz, *Imposter Syndrome and The 'As-If' Personality in Analytical Psychology: The Fragility of Self* (Taylor & Francis, 2023), <https://doi.org/10.4324/9781003315254>.

20. K. A. Phillips, "Body Dysmorphic Disorder," in *Clinical Obsessive-Compulsive Disorders in Adults and Children* (Cambridge University Press, 2011), 191, <https://doi.org/10.1017/CBO9780511973796.015>.
21. C. Pittenger, B. P. Brennan, L. Koran, et al., "Specialty Knowledge and Competency Standards for Pharmacotherapy for Adult Obsessive-Compulsive Disorder," *Psychiatry research* 300 (2021): 113853, <https://doi.org/10.1016/j.psychres.2021.113853>.
22. G. Trundle, and R. Hutchinson, "The Phased Model of Adventure Therapy: Trauma-Focussed, Low Arousal, & Positive Behavioural Support," *Journal of Adventure Education and Outdoor Learning* 21, no. 1 (2021): 82–92, <https://doi.org/10.1080/14729679.2020.1736109>.
23. M. B. Cororve, and D. H. Gleaves, "Body Dysmorphic Disorder: A Review of Conceptualizations, Assessment, and Treatment Strategies," in *Obsessive-Compulsive Disorder and Tourette's Syndrome* (Routledge, 2022), 13–34, <https://doi.org/10.4324/9780203822937-2>.
24. D. Veale, L. J. Gledhill, P. Christodoulou, and J. Hodsoll, "Body Dysmorphic Disorder in Different Settings: A Systematic Review and Estimated Weighted Prevalence," *Body Image* 18 (2016): 168–186, <https://doi.org/10.1016/j.bodyim.2016.07.003>.
25. M. S. Alsaidan, N. S. Altayar, S. H. Alshmmari, M. M. Alshammari, F. T. Alqahtani, and K. A. Mohajer, "The Prevalence and Determinants of Body Dysmorphic Disorder Among Young Social Media Users: A Cross-Sectional Study," *Dermatology Reports* 12, no. 3 (2020), <https://doi.org/10.4081/dr.2020.8774>.
26. L. Schreurs, and L. Vandenbosch, "Different Interactions With Appearance-Focused Social Media Content and Adolescents' Body Dissatisfaction: A Within-Person Perspective," *Computers in Human Behavior* 135 (2022): 107364, <https://doi.org/10.1016/j.chb.2022.107364>.
27. L. R. McGrath, L. Oey, S. McDonald, D. Berle, and B. M. Wootton, "Prevalence of Body Dysmorphic Disorder: A Systematic Review and Meta-Analysis," *Body Image* 46 (2023): 202–211, <https://doi.org/10.1016/j.bodyim.2023.06.008>.
28. G. Krebs, L. F. de la Cruz, and D. Mataix-Cols, "Recent Advances in Understanding and Managing Body Dysmorphic Disorder," *BMJ Mental Health* 20, no. 3 (2017): 71–75, <https://doi.org/10.1136/eb-2017-102702>.
29. S. C. Schneider, J. Mond, C. M. Turner, and J. L. Hudson, "Subthreshold Body Dysmorphic Disorder in Adolescents: Prevalence and Impact," *Psychiatry Research* 251 (2017): 125–130, <https://doi.org/10.1016/j.psychres.2017.01.085>.
30. A. R. Singh, and D. Veale, "Understanding and Treating Body Dysmorphic Disorder," Supplement 1, *Indian Journal of Psychiatry* 61 (2019): S131–S135, https://doi.org/10.4103/psychiatry.IndianJPsychiatry_528_18.
31. R. F. Hakim, D. A. Alrahmani, D. M. Ahmed, N. A. Alharthi, A. R. Fida, and R. M. Al-Raddadi, "Association of Body Dysmorphic Disorder With Anxiety, Depression, and Stress Among University Students," *Journal of Taibah University Medical Sciences* 16, no. 5 (2021): 689–694, <https://doi.org/10.1016/j.jtumed.2021.05.008>.
32. C. G. Santos, M. Almeida, M. L. de Oliveira Junior, T. A. Brown, and P. H. B. de Carvalho, "Psychometric Evaluation of the Drive for Muscularity Scale and the Muscle Dysmorphic Disorder Inventory Among Brazilian Cisgender Gay and Bisexual Adult Men," *International Journal of Environmental Research and Public Health* 20, no. 2 (2023): 989, <https://doi.org/10.3390/ijerph20020989>.
33. F. Incitti, F. Urli, and L. Snidaro, "Beyond Word Embeddings: A Survey," *Information Fusion* 89 (2023): 418–436, <https://doi.org/10.1016/j.inffus.2022.08.024>.
34. T. Theodoropoulos, A. Makris, I. Kontopoulos, et al., "Graph Neural Networks for Representing Multivariate Resource Usage: A Multiplayer Mobile Gaming Case-Study," *International Journal of Information Management Data Insights* 3, no. 1 (2023): 100158, <https://doi.org/10.1016/j.jjime.2023.100158>.
35. S. L. King, and T. Neal, "Applications of AI-Enabled Deception Detection Using Video, Audio, and Physiological Data: A Systematic Review," *IEEE Access* 12 (2024): 135207–135240, <https://doi.org/10.1109/ACCESS.2024.3462825>.
36. Y. Huang, J. Xu, J. Lai, et al., "Advancing Transformer Architecture in Long-Context Large Language Models: A Comprehensive Survey," <https://arxiv.org/abs/2311.12351>.
37. D. Kodati, and R. Tene, "Advancing Mental Health Detection in Texts via Multi-Task Learning With Soft-Parameter Sharing Transformers," *Neural Computing and Applications* 37 (2025): 3077–3110, <https://doi.org/10.1007/s00521-024-10753-7>.
38. A. Khamaj, and A. M. Ali, "Adapting User Experience With Reinforcement Learning: Personalizing Interfaces Based on User Behavior Analysis in Real-Time," *Alexandria Engineering Journal* 95 (2024): 164–173, <https://doi.org/10.1016/j.aej.2024.03.045>.
39. T. Hussain, L. Yu, M. Asim, A. Ahmed, and M. A. Wani, "Enhancing E-Learning Adaptability With Automated Learning Style Identification and Sentiment Analysis: A Hybrid Deep Learning Approach for Smart Education," *Information* 15, no. 5 (2024): 277, <https://doi.org/10.3390/info15050277>.
40. A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current State of Text Sentiment Analysis From Opinion to Emotion Mining," *ACM Computing Surveys (CSUR)* 50, no. 2 (2018): 1–33, <https://doi.org/10.1145/3057270>.
41. C. Dhaoui, C. M. Webster, and L. P. Tan, "Social Media Sentiment Analysis: Lexicon Versus Machine Learning," *Journal of Consumer Marketing* 34, no. 6 (2017): 480–488, <https://doi.org/10.1108/JCM-03-2017-2141>.
42. R. A. Calvo, and S. Mac Kim, "Emotions in Text: Dimensional and Categorical Models," *Computational Intelligence* 29, no. 3 (2013): 527–543, <https://doi.org/10.1111/j.1467-8640.2012.00456.x>.
43. K. Setkowski, K. Boogert, A. W. Hoogendoorn, R. Gilissen, and A. J. van Balkom, "Guidelines Improve Patient Outcomes in Specialised Mental Health Care: A Systematic Review and Meta-Analysis," *Acta Psychiatrica Scandinavica* 144, no. 3 (2021): 246–258, <https://doi.org/10.1111/acps.13332>.