JMIR Serious Games

Journal Impact Factor (JIF) (2023): 3.8
Volume 14 (2026) ISSN 2291-9279 Editor-in-Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Corrigenda and Addenda

Research Letter

Usability Study of Augmented Reality Visualization Modalities on Localization Accuracy in the Head and Neck: Randomized Crossover Trial

Yao Li[1,2], MSc; Gijs Luijten[3,4,5], MSc; Christina Gsaxner[3,5], PhD; Kim Grunert[1,2], MD; Alexis Bader[1,2,3,6]; Frank Hölzle[1], MD, DMD, PhD; Rainer Röhrig[2], MD, PhD; Matías de la Fuente[7], PhD; Jan Egger[3,4], PhD; Kunpeng Xie[1,2], MSM; Behrus Hinrichs-Puladi[1,2], MD, DMD

[1]Department of Oral and Maxillofacial Surgery, University Hospital RWTH Aachen, Pauwelsstrasse 30, Aachen, Germany
[2]Institute of Medical Informatics, University Hospital RWTH Aachen, Aachen, Germany
[3]Institute for Artificial Intelligence in Medicine (IKIM), Essen University Hospital, Essen, Germany
[4]Center for Virtual and Extended Reality in Medicine (ZvRM), Essen University Hospital, Essen, Germany
[5]Institute of Computer Graphics and Vision (ICG), Graz University of Technology, Graz, Austria
[6]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada
[7]Chair of Medical Engineering, RWTH Aachen University, Aachen, Germany

Corresponding Author:
Kunpeng Xie, MSM
Department of Oral and Maxillofacial Surgery, University Hospital RWTH Aachen, Pauwelsstrasse 30, Aachen, Germany

Abstract

Background: Augmented reality head-mounted displays could overcome the spatial dissociation between medical imaging and the surgical field, which may be particularly important in anatomically dense regions, such as the head and neck. Although many head-mounted displays offer markerless inside-out tracking at a fraction of the cost of navigation systems, their overlay accuracy with superimposition (SI) modality onto the surgical field remains limited. The virtual twin (VT), displaying holography adjacent to the surgical field, may offer a viable alternative. However, its performance is still unclear.

Objective: This study aimed to compare the accuracy and efficiency of the two visualization modalities, SI and VT, for anatomical localization in the head and neck region.

Methods: In a randomized crossover trial to compare two augmented reality visualization modalities (SI and VT), 38 participants used a HoloLens 2 to localize point, line-based, and volume-based anatomical structures on head phantoms. Their performance was evaluated with respect to accuracy, workload, time, and user experience.

Results: SI achieved significantly better point localization accuracy than VT both in absolute (mean 14.4, SD 4.2 mm vs mean 15.8, SD 5.5 mm; P=.003) and relative accuracy (mean 3.4, SD 2.2 mm vs mean 6.0, SD 5.0 mm; P<.001). In line-based structures, accuracy was comparable between SI (average surface distance [ASD], mean 23.4, SD 4.1 mm; Hausdorff distance [HD], mean 31.5, SD 7.8 mm) and VT (ASD=mean 23.0, SD 4.5 mm; P=.51; HD=mean 31.0, SD 7.5 mm; P=.57). However, SI showed significantly higher deviation than VT in volume-based structure (ASD=mean 37.1, SD 13.8 mm vs mean 34.1, SD 14.2 mm; P=.01; HD=mean 52.0, SD 16.8 mm vs mean 49.1, SD 15.8 mm; P=.03). Participants were faster with SI (P=.02), while workload NASA-TLX (National Aeronautics and Space Administration Task Load Index) scores did not demonstrate a significant difference (P=.79).

Conclusions: Given that SI did not clearly outperform VT under overlaid soft tissue and viewing challenges, VT remains a viable alternative in certain surgical scenarios where high accuracy is not required. Future research should focus on optimizing viewing angle guidance and the linkage between the anatomical target and the skin surface.

Trial Registration: German Clinical Trial Register DRKS00032835; https://drks.de/search/en/trial/DRKS00032835

(JMIR Serious Games 2026;14:e75962) doi:10.2196/75962

KEYWORDS

mixed reality; computer-assisted surgery; visualization techniques; human-machine interface; preoperative planning

Introduction

The head and neck region contains a variety of complex anatomical structures, including numerous vital nerves, blood vessels, and organs [1]. Accurate localization of these anatomical structures is crucial in surgical practice to minimize deviation and improve outcomes [2]. Conventional medical imaging techniques, such as computed tomography (CT) and cone beam CT, as well as magnetic resonance imaging, are primarily used for diagnosis and preoperative planning [3,4]. Medical images require surgeons to mentally map medical images onto the patient's anatomy during the operation. This process demands a high level of cognitive effort, especially in the anatomically dense head and neck region, where misinterpretation could compromise the surgical accuracy and outcomes [5,6]. Surgical navigation systems (SNS) offer solutions by integrating image data into the surgical workflow. However, the limitations of the 3D display still leave the operator reliant on spatial imagination to understand complex anatomy. Furthermore, the broader adoption of SNS has been impeded by high expenses, the inherently sophisticated configurations like optical tracking cameras and reflective markers, and the possible additional radiation exposure to patients and staff [7,8]. As a result, there is still a lack of a cost-effective, intuitive, 3D interactive visualization approach that seamlessly displays the patient's medical images in the field.

Augmented reality (AR) could fill this gap by providing real-time holographic images directly within the surgical field mainly through head-mounted displays (HMDs) [7,9]. Moreover, many current AR HMDs can provide markerless inside-out tracking at a fraction of the cost of SNS and eliminate the need for additional markers [10,11]. Unlike SNS, which typically tracks the patient and instruments, this kind of HMD-based tracking focuses on aligning virtual content with the patient's anatomy to enable hologram overlay. However, the overlay or registration accuracy of many HMDs is still not as accurate as traditional SNSs with external optical tracking at the millimeter level [10]. This limitation becomes particularly critical for the superimposition (SI) visualization modality, where virtual anatomical structures need to be precisely placed on real anatomy, a process referred to as registration [12-14]. In addition, SI may introduce occlusion, as holograms can obstruct the surgeon's view of anatomy or instruments. These challenges raise concerns about the feasibility of SI as the optimal visualization modality for AR-assisted surgery, given the setup of currently available HMDs free of external tracking [15].

An alternative visualization modality is the virtual twin (VT), where the holographic representation is displayed adjacent to the physical anatomy instead of directly overlaid on the anatomy [15,16]. By avoiding overlay, VT reduces dependence on registration accuracy and eliminates occlusion.

However, the accuracy between two modalities under markerless HMD-based tracking remains unexplored. Yet, this could be important, since if SI with intrinsic markerless tracking does not show any advantage over VT, then VT would be the favored modality for certain surgical scenarios. Therefore, the aim of this crossover randomized controlled trial (RCT) was to compare the accuracy and efficiency of the two visualization modalities, SI and VT, for anatomical localization in the head and neck region. Localization accuracy was assessed on phantom heads for clinically relevant targets, including nerve exit points, the inferior alveolar nerve, and the salivary glands. Task duration and subjective workload were evaluated as secondary endpoints.

Methods

Overview

In total, 38 participants with different professional backgrounds (dental and medical students, resident and specialist surgeons in oral and maxillofacial, oral, and plastic surgery) were recruited and performed drawings on polystyrene foam head phantoms (Model SAM, Friseurbedarf D. M. Rudolph) in a crossover RCT with SI and VT visualization modalities. The participants were asked to draw the structures on the head phantoms, wearing HoloLens 2 (HL2; Microsoft Corp). The primary endpoint was the localization accuracy of the anatomical points (0D), which encompass nerve exit points at the supraorbital, infraorbital, and mental foramina. Secondary endpoints included the delineation accuracy of the inferior alveolar nerve pathways (2D) and salivary glands (parotid and submandibular; 3D), cognitive workload, and user experience.

System Description and Implementation

The AR visualization software for the HL2 was developed in-house to display anatomical 3D models in relation to the physical anatomy of patients or phantoms. Within the application, switching between the two different visualization modalities for the 3D models was possible. In addition to the HL2 software, a pipeline processed the medical image data. This pipeline converted volumetric CT scans into 3D models optimized for interventional planning and efficient rendering on the HL2.

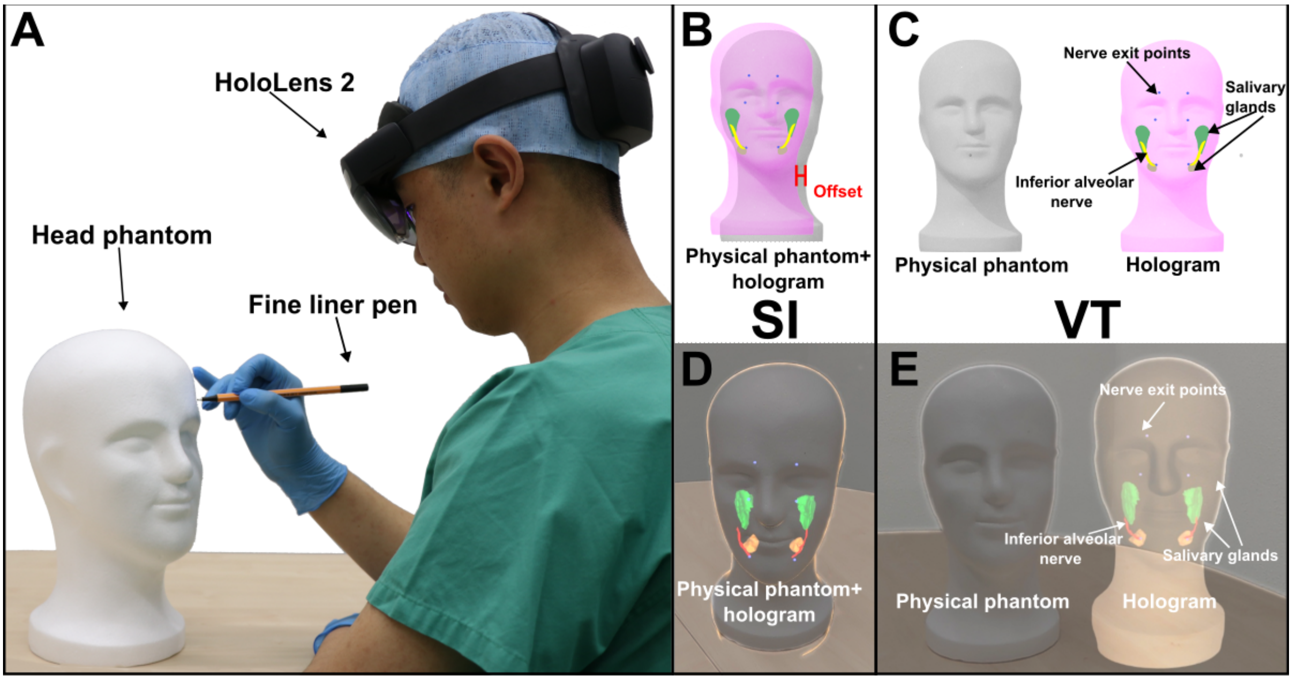
Based on these requirements, the planning pipeline was built to segment the structures into meshes in 3D Slicer (version 5.2.2; The Slicer Community). The structures comprising the skull, salivary glands, and nerve exit points were manually segmented from a publicly available head and neck CT dataset [17], while the inferior alveolar nerves were segmented from a nonpublic dataset from the Medical University of Graz. A head phantom mesh was scanned by the Artec Leo 3D scanner (Artec 3D) as the skin surface. Finally, all segmented anatomical structures were nonlinearly registered to the scanned skin surface.

Our AR application was developed using Unity (version 2022.3.6f1; Unity Technologies). The registration between the head phantoms and the virtual head was implemented using the Vuforia software development kit (version 10.16.5, Parametric Technology Corporation). Vuforia Engine is a cross-platform AR solution that offers a variety of tracking features, which was frequently used in research for AR registration in surgical scenarios [18-20]. The model targets (object tracking) were applied, which possibly used edge-based techniques (not revealed by Vuforia) to recognize and track objects in real-time [21]. First, the scanned head model was uploaded to the model target generator tool and configured into a model target that could be integrated into Unity. After the software was deployed to the HL2, the Vuforia engine initiated tracking for target alignment. Once the participant is satisfied with the alignment, she or he could lock the tracking to anchor the virtual model in the environment. Similarly, in VT, Vuforia would track the phantom, and then the model would appear next to it; locking the tracking again would fix the model in place. The hand menu assisted users in controlling the visibility of various anatomical structures, including the skin and target structures. In addition, sliders were implemented to allow real-time adjustment of the transparency and brightness of these structures.

Trial

The participants were asked to fill out the initial questionnaire, which included demographic information (age, gender, educational stage or professional experience, professional field, and prior experience with AR and HL2). Randomization was generated by BHP using a randomized allocation rule to determine the starting modality (sequence) and the side of the face (right or left). The experiment assistant (KG) enrolled and assigned participants to the sequence of intervention. Registration was done once at the beginning of each modality by the experiment assistant, who could lock or unlock the tracking for registration as needed. Subsequently, they wore HL2, ran the eye calibration, and received a brief introduction to the device and the user interface with the 2 modalities. During this short session, they familiarized themselves with the device and its functions. The entire familiarization process was completed in less than 3 minutes, although precise timing was not recorded. Participants were then instructed to delineate target anatomical structures on the head phantom surface using Point 88 fine liner pens (Stabilo; Figure 1). This task was performed on the assigned half of the face using the first modality, with time recorded via a stopwatch. Upon completion, participants filled out the Likert questionnaire and NASA-TLX (National Aeronautics and Space Administration Task Load Index) for that method and an open-ended questionnaire. The same procedure was then repeated on the other side of the face using the second modality, followed by the corresponding questionnaires. Finally, an open-ended questionnaire for preference was answered.

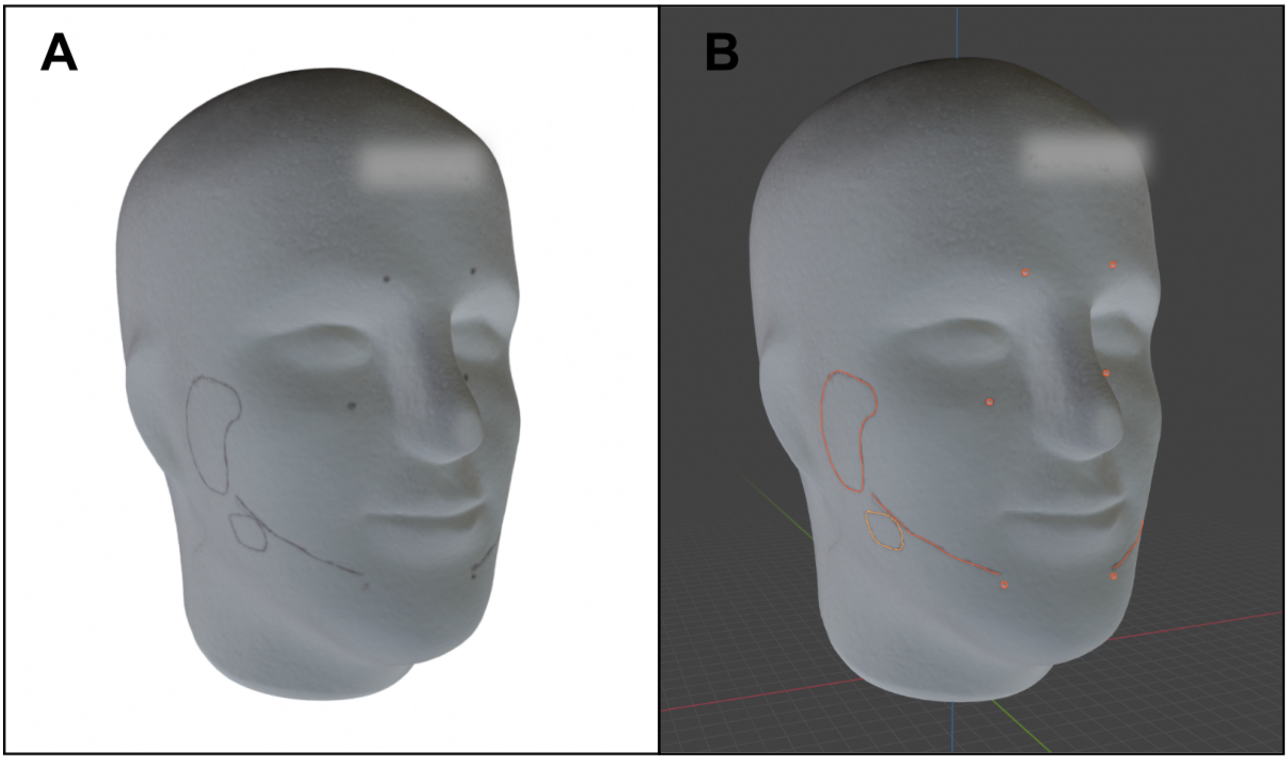
Figure 1. Illustration of two augmented reality visualization modalities using HoloLens 2. (A) Participant drawing anatomical structures (nerve exit points, inferior alveolar nerves, and salivary glands) on the polystyrene head phantom with HoloLens 2. (B) Schematic illustration of SI showing physical and holographic alignment with potential rigid offset and occlusion. (C) Schematic illustration of virtual twin showing how holograms are displayed free of misalignment and occlusion problem. (D) SI modality in HoloLens 2, where holograms were overlaid directly into the physical head phantom. (E) Virtual twin modality in HoloLens 2, where the holograms were displayed spatially adjacent to the physical head phantom. SI: superimposition; VT: virtual twin.



Evaluation

After the trial, all the polystyrene head phantoms were scanned with the Artec Leo 3D scanner (Figure 2). To enable comparison, all head phantoms with the participant's delineations were registered to the virtual planned head in a pipeline by a Python (version 3.10; Python Software Foundation) script. The two-stage pipeline was initiated with a global random sample consensus alignment, followed by a local refinement with point-to-plane iterative closest point, achieving <0.4 mm root mean square error. Two independent investigators (YL and KG) evaluated the scanned heads using Blender (version 4.2; Blender Foundation). Both investigators were blinded to the applied visualization modality. To minimize a possible recall bias, KG, who served as the experiment assistant during data acquisition, underwent a washout period of 2 months before participating in the blinded evaluation. Nerve exit points were drawn by the stroke points and placed spheres. The nerve paths and salivary glands were drawn by the grease pencil tool along the curves on the head phantom surface, and the strokes were transformed into meshes in Blender (Figure 2).

Figure 2. (A) Scanned polystyrene head phantom with delineation. (B) In Blender, the scanned polystyrene head phantom is shown with inferior alveolar nerve and salivary glands annotations, and spheres marking the nerve exit points (orange).



Afterward, the points and curves were automatically compared using a Python script. The analysis for nerve exit points (0D) involved calculating Euclidean distance, which is the shortest distance in 3D space between the planned and drawn points, and we referred to this as absolute accuracy. Relative accuracy, defined as landmark-to-landmark localization accuracy, was compared by the Euclidean distance between the supraorbital-infraorbital and infraorbital-mental foramina on drawn versus planned landmarks. Since all anatomical targets were located on the underlying bone, yet localization was performed on the phantom's external surface, the concept of soft-tissue thickness was additionally introduced to capture the distance between the target structures and the skin. It was defined as the shortest distance from each anatomical point (0D) to the surface and as the mean of the vertex-to-surface distances for 2D nerve pathways and 3D salivary glands. Furthermore, the Hausdorff distance (HD) and the average surface distance (ASD) were used in order to assess the alignment and accuracy of the contours of the nerve paths (line, 2D) and salivary glands (volume, 3D). HD captures the maximum of the minimum distances between the two surfaces, providing insight into the worst-case alignment error, while the ASD quantifies the mean discrepancy, reflecting the overall degree of alignment.

The Likert questionnaire and NASA-TLX were quantitatively analyzed to assess usability and perceived workload. In addition, the feedback from open-ended questions was summarized by YL and reviewed by BHP.

Sample Size Calculation

The sample size calculation was conducted in R software (version 4.3.1; R Foundation for Statistical Computing). A minimum effect size of 5 mm was established as the threshold for an acceptable difference between the two modalities in absolute accuracy. A 5 mm difference in absolute accuracy causes a surface discrepancy exceeding 5 mm due to the geometric relationship, making it clinically relevant and detectable by oral and maxillofacial surgeons, corresponding to the widely accepted minimum margin in head and neck oncologic surgery [22,23]. Based on the results of a pretrial with 4 participants, the mean absolute accuracy was 10.1 (SD 4.8) mm (SI) and 12.1 (SD 5.0) mm (VT) across all nerve exit points. A normal distribution of the pretest values (Shapiro-Wilk test; *P*=.70) resulted in a required number of cases of 34 for the unpaired *t* test. An additional 4 participants were included to compensate for nonevaluable datasets and for dropout or withdrawal of consent.

Statistical Analysis

Statistical analysis was also performed in R. A linear mixed-effects model (LMM) was applied using the *lmerTest* package [24]. This LMM assessed the absolute accuracy at the point structures, modalities (SI vs VT), the sequence (starting method), the group (dental and medical students, and surgeons), subcutaneous soft tissue thickness, and side (left or right) as fixed effects and the participants as a random effect. When analyzing the ASD and HD for line and volume-based structures, the same LMM framework was applied. Subcutaneous soft tissue thickness was specifically included to account for anatomical variation across different locations. However, it was not considered in the analysis of relative accuracy for point structures, which instead relied more on spatial reference to other anatomical landmarks.

The normality of the data distribution was assessed using the Shapiro-Wilk test. Duration and each Likert question between methods were compared using the Mann-Whitney *U* test. The NASA-TLX scales were compared by unpaired two-tailed *t* test. For all tests mentioned, a *P* value of <.05 was considered significant.

Ethical Considerations

This study was approved by the local ethics committee of the University Hospital RWTH (Rheinisch-Westfälische Technische Hochschule) Aachen (EK 24‑127; Chairman Prof Ralf Hausmann; April 3, 2024). The study was registered with a study protocol in advance in the German Clinical Trial Register (DRKS00032835) and followed the CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines (Checklist 1) and its extension designed and modified specifically for crossover studies, as illustrated by the flow diagram [25,26]. Informed consent was obtained from all participants involved in the study. To protect the privacy of the participants, all participants were anonymized, and no personally identifiable information was stored with the research data. It is to be noted that no financial compensation was provided to the participants involved in the present trial. Nevertheless, as a token of appreciation, two vouchers with a total value of €15 (US $17.5) were distributed through a raffle.

Results

Cohort

A total of 38 participants (16 females and 22 males) were successfully included in the study, comprising two groups, namely surgeons, and medical and dental students following the flow (Figure 3). Among the 18 surgeons, there were 12 residents and 6 specialists. This group included 9 oral and maxillofacial surgeons, 5 oral surgeons, and 4 plastic surgeons. In the student group, which consisted of 20 participants, 17 were dental students and 3 were medical students. The average age of participants was 26.8 (SD 5.1; range 20‑43). The average clinical experience of surgeons was 4.0 (SD 4.3) years, and the average clinical experience of medical and dental students was 4.2 (SD 0.9) years; mean 8.3 (SD 1.6) semesters (Table 1).

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) flow diagram illustrating the enrollment, allocation, crossover, follow-up, and analysis of participants in the study. SI: superimposition; VT: virtual twin.

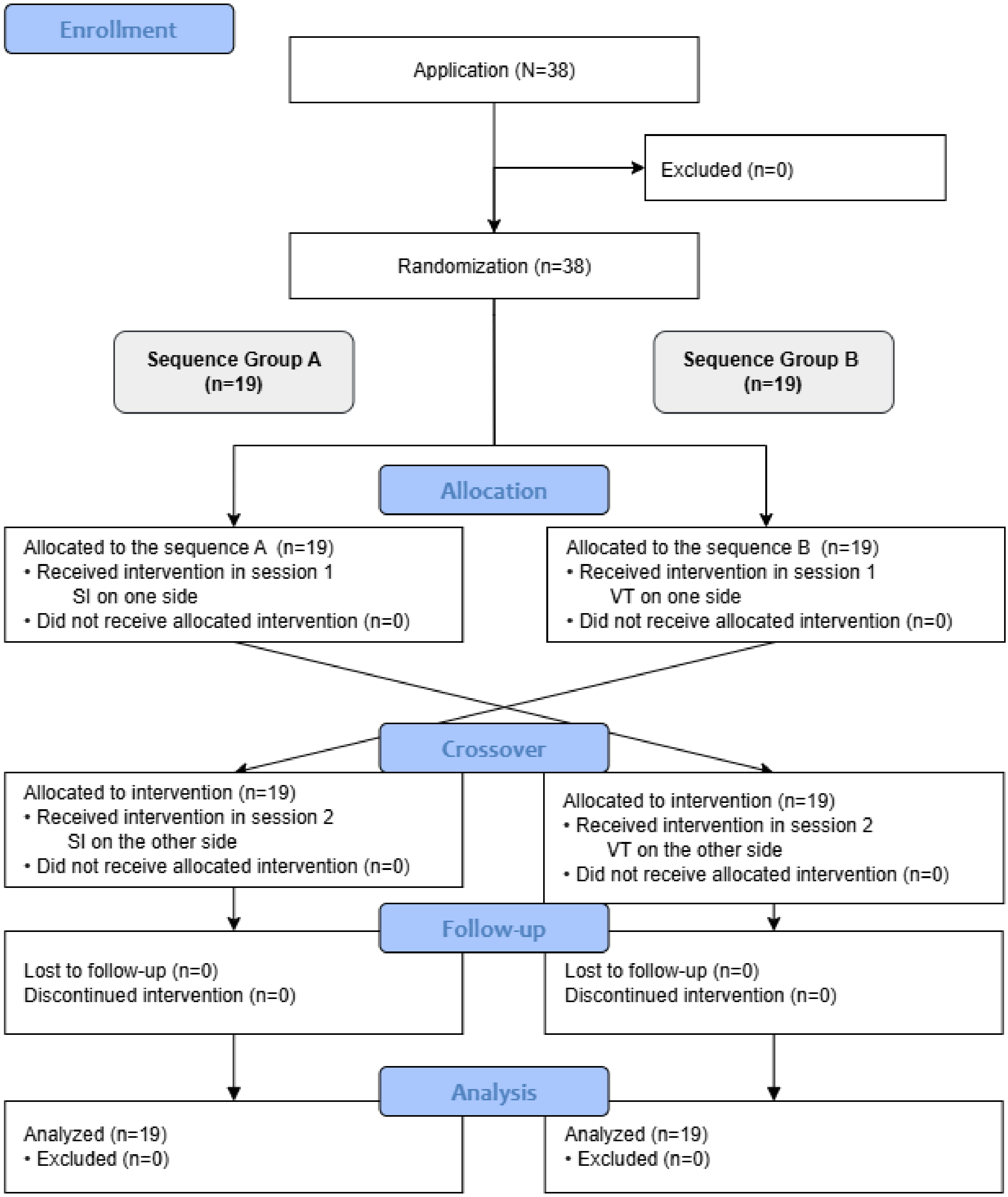


Table . Characteristics of the cohort.

Parameter	Surgeon (n=18)	Student (n=20)	Total (n=38)
Sex, n (%)			
Female	4 (22.2)	12 (60)	16 (42.1)
Male	14 (77.8)	8 (40)	22 (57.9)
Age (years)			
Mean (SD)	30.3 (5.2)	23.8 (2.5)	26.8 (5.1)
Range	23-43	20-30	20-43
Profession, n (%)			
Medical	—[a]	3 (15)	3 (7.9)
Dental student	—	17 (85)	17 (44.7)
Oral surgery	5 (27.8)	—	5 (13.2)
Oral and maxillofacial surgery	9 (50)	—	9 (23.7)
Plastic surgery	4 (22.2)	—	4 (10.5)
Clinical study/work experience (years)			
Mean (SD)	4.0 (4.3)	4.2 (0.9)	4.1 (3.0)
Range	0.0-15.0	3.0-6.0	0.0-15.0
Previous experience with AR (Likert score, 1–5)[b]			
Mean (SD)	2.4 (0.8)	2.1 (1.0)	2.2 (0.9)
Range	1.0-4.0	1.0-5.0	1.0-5.0
Previous experience with HL (Likert score, 1–5)[c]			
Mean (SD)	1.7 (0.8)	1.1 (0.3)	1.4 (0.6)
Range	1.0-3.0	1.0-2.0	1.0-3.0

[a]Not applicable.

[b]AR: augmented reality; Likert scores from 1=”never heard of” to 5=”expert.”

[c]HL: HoloLens; Likert scores from 1=“never used” to 5=“I use it several times a week.”

Localization Accuracy

In the 38 scanned head phantoms, all the required structures were successfully delineated, except for 1 pair of nerve exit points at infraorbital foramina and 1 pair at supraorbital foramina, which were missed by a single participant. The absolute accuracy of the nerve exit points (0D) was significantly higher in SI (mean 14.4, SD 4.2 mm) than VT (mean 15.8, SD 5.5 mm), with a mean difference of 1.4 (95% CI 0.5 - 2.3; LMM; *P*=.003) mm. The absolute accuracy was correlated with the soft tissue thickness. For each 1 mm soft tissue thickness, the accuracy decreased by 1.4 mm (*P*<.001), while no significant difference was found in sequence (*P*=.84) and group (*P*=.40) as fixed effects in the LMM. The average participant bias was 0.8 (SD 0.8) mm. The mean absolute error of the LMM residuals was 1.8 (SD 2.9) mm for SI and 2.5 (SD 3.5) mm for VT, respectively. The relative accuracy of the points was significantly higher for SI (mean 3.4, SD 2.2 mm) than VT (mean 6.0, SD 5.0 mm) by 2.6 (95% CI 1.3 - 3.8 mm; LMM; *P*<.001; Figure 4). In Figure 4, each violin plot (colored) includes a boxplot (white), with a red dot indicating the mean value. The black points represent the outliers. The dashed line marked the average subcutaneous soft tissue thickness over the nerve exit points.

The localization accuracy of the inferior alveolar nerve pathways (2D) assessed with ASD and HD was comparable between SI (ASD/HD=mean 23.4, SD 4.1 mm/mean 31.5, SD 7.8 mm) and VT (ASD/HD=mean 23.0, SD 4.5 mm/mean 31.0, SD 7.5 mm), with no significant difference (ASD/HD=mean difference 0.4 mm, 95% CI −1.0 to 2.0 mm; LMM; *P*=.51/mean difference 0.6 mm, 95% CI –1.6 to 2.9 mm; LMM; *P*=.57). Regarding the salivary glands (3D), the localization accuracy measured with ASD/HD (mean 34.1, SD 14.2 mm/mean 49.1, SD 15.8 mm) for VT was significantly more accurate than SI (ASD/HD=mean 37.1, SD 13.8 mm/mean 52.0, SD 16.8 mm) by ASD 3.0 (95% CI 0.7 - 5.4 mm; LMM; *P*=.01) mm and HD 2.9 (95% CI 0.2 - 5.8 mm; LMM; *P*=.03) mm (Figure 5). In Figure 5, each violin plot (colored) includes a boxplot (white), with a red dot indicating the mean value. The black points represent the outliers. The dashed line marked the average subcutaneous soft tissue thickness over the inferior alveolar nerves and salivary glands.

Figure 4. Comparison of localization accuracy (y-axis) at nerve exit points (0D) between superimposition (purple) and virtual twin (green; x-axis). (A) Euclidean distance for absolute accuracy. (B) Absolute residual error from the linear mixed-effects model. (C) Euclidean distance for relative accuracy. (D) Relationship between Euclidean distance for absolute accuracy (y-axis) and subcutaneous soft tissue thickness (x-axis). The solid blue line depicts the fitted linear mixed-effects model regression. ED: Euclidean distance; LMM: linear mixed-effects model; SI: superimposition; VT: virtual twin.

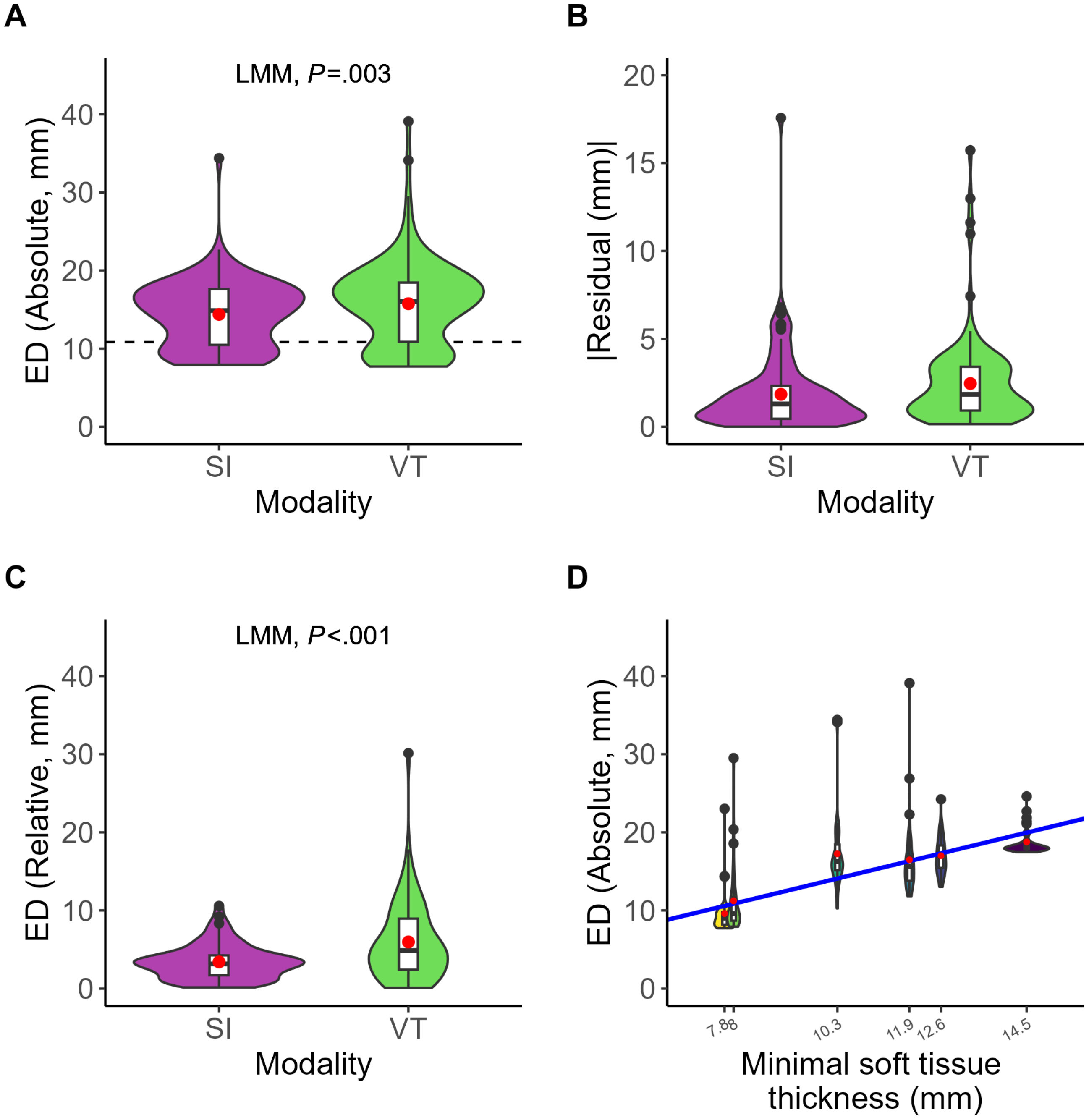
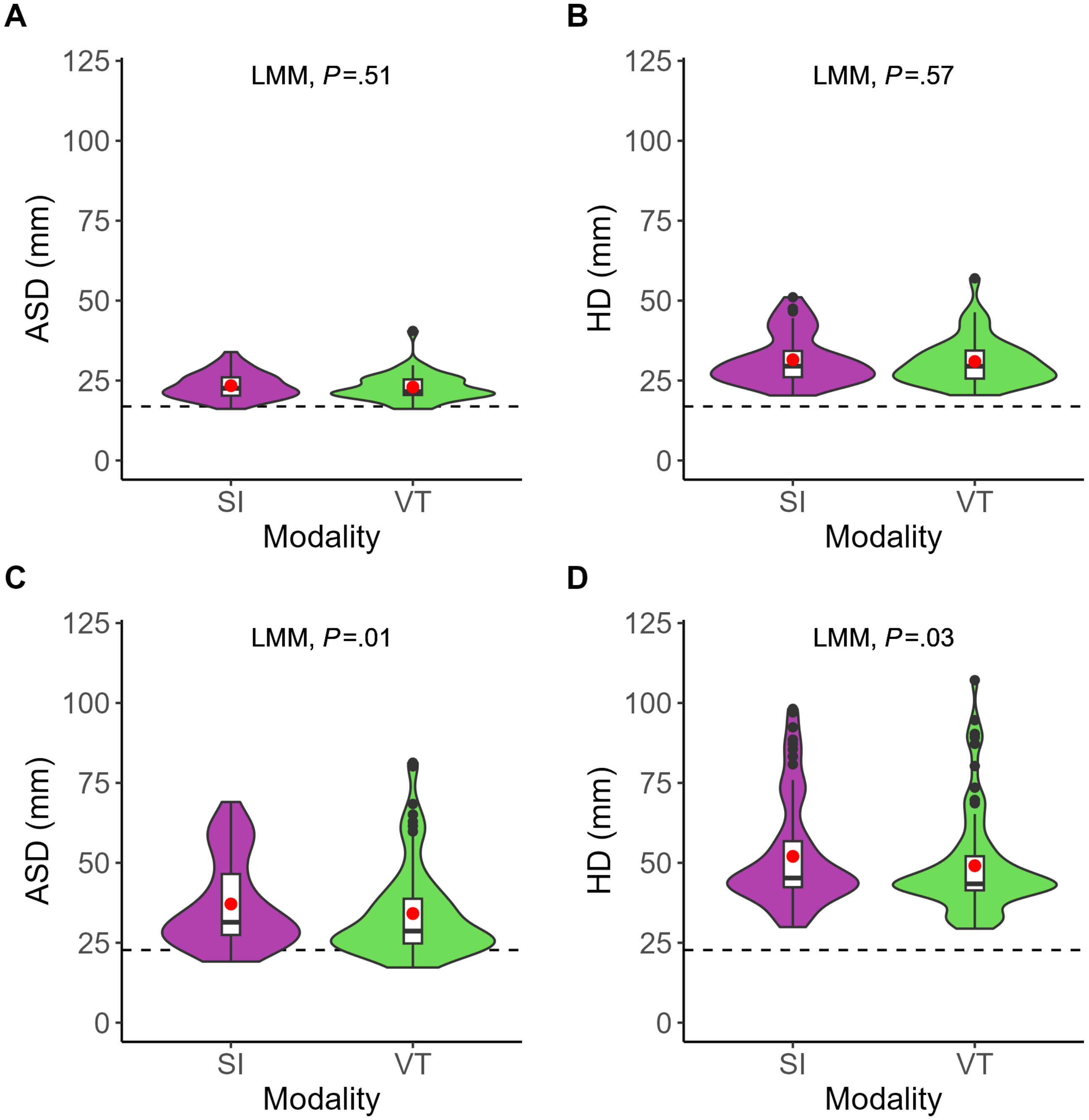


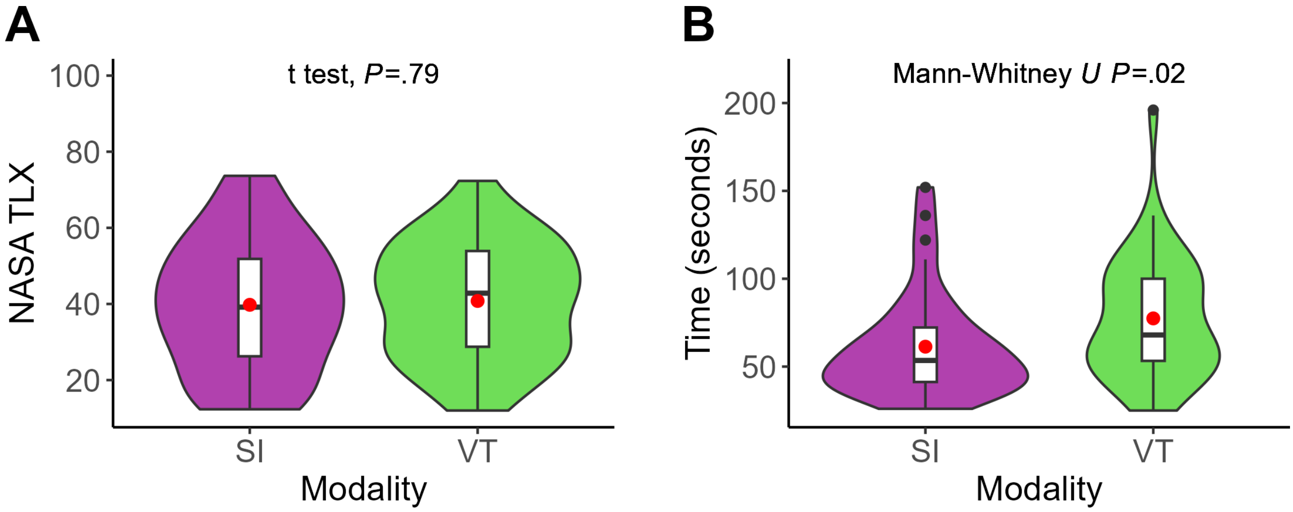
Figure 5. Comparison of localization accuracy (y-axis) for inferior alveolar nerve pathways (2D) and salivary glands (3D) between superimposition (purple) and virtual twin (green; x-axis). (A) Average surface distance for inferior alveolar nerve pathways (2D). (B) Hausdorff distance for inferior alveolar nerve pathways (2D). (C) Average surface distance for salivary glands (3D). (D) Hausdorff distance for salivary glands (3D). ASD: average surface distance; HD: Hausdorff distance; LMM: linear mixed-effects model; SI: superimposition; VT: virtual twin.



Workload and Time

The SI method (mean 61.3, SD 29.6 seconds) was significantly faster than the VT method (mean 77.4, SD 34.5 seconds) by 16.1 (95% CI 2.0 - 29.0; Mann-Whitney *U* test; *P*=.02) seconds. The NASA-TLX score for the SI method (mean 39.8, SD 17.3) and VT method (mean 40.8, SD 15.2) was comparable, with no significant difference (mean difference 1.0, 95% CI –4.2 to 6.2; *t* test; *P*=.79; Figure 6). In Figure 6, each violin plot (colored) includes a boxplot (white), with a red dot indicating the mean value. The black points represent the outliers.

Figure 6. Subjective ratings and task completion time between superimposition (purple) and virtual twin (green) visualizations (x-axis). (A) Subjective workload assessed using NASA-TLX (National Aeronautics and Space Administration Task Load Index) scores (y-axis). (B) Task completion time in seconds (y-axis). NASA TLX: National Aeronautics and Space Administration Task Load Index; SI: superimposition; VT: virtual twin.



Questionnaires

The Likert-type questions (scale 1-4; 1=strong disagreement; 4=strong agreement) showed no significant difference (Mann-Whitney *U* test) between the two modalities (Table 2). The participants perceived no clear advantage in accurate localization of target structures between SI and VT (mean 3.0, SD 0.9 vs mean 3.0, SD 0.6 points; *P*=.61; mean 2.9, SD 0.8 vs mean 2.8, SD 0.7 lines; *P*=.37; mean 2.7, SD 0.8 vs mean 2.7, SD 0.7 volume; *P*=.95). Participants also reported similar levels of confidence (mean 2.7, SD 0.7 vs mean 2.7, SD 0.6; *P*=.84), distraction (mean 2.2, SD 1.0 vs mean 1.8, SD 0.9; *P*=.05), provided assistance (mean 2.9, SD 0.8 vs mean 3.1, SD 0.6; *P*=.46), practicality (mean 2.4, SD 0.9 vs mean 2.8, SD 0.8; *P*=.09), perceived feasibility in interventions (mean 2.8, SD 1.1 vs mean 2.7, SD 0.9; *P*=.57), safety enhancement (mean 2.5, SD 0.9 vs mean 2.7, SD 0.9; *P*=.19), and overall satisfaction (mean 2.8, SD 0.9 vs mean 2.9, SD 0.8; *P*=.66). In addition, positive and negative detailed feedback was provided for both visualization modalities (Table 3). It is noteworthy that 19 participants expressed a preference for VT, 18 participants for SI, and 1 participant expressed equal preference for both.

Table . Likert questionnaire.

Likert questions	SI[a], mean (SD)	VT[b], mean (SD)	Total, mean (SD)	*P* value
I was able to accurately mark the nerve exit points using the (SI or VT) visualization.	3.0 (0.9)	3.0 (0.6)	3.0 (0.8)	.61
I was able to accurately mark the nerve pathways using the (SI or VT) visualization.	2.9 (0.8)	2.8 (0.7)	2.8 (0.7)	.37
I was able to accurately mark the salivary glands using the (SI or VT) visualization.	2.7 (0.8)	2.7 (0.7)	2.7 (0.7)	.95
I was sure where the anatomical structures were located and where to mark them.	2.7 (0.7)	2.7 (0.6)	2.7 (0.7)	.84
I found the using (SI or VT) visualization distracting while marking.	2.2 (1.0)	1.8 (0.9)	2.0 (1.0)	.05
The using (SI or VT) visualization facilitated the localization of anatomical structures in the face.	2.9 (0.8)	3.1 (0.6)	3.0 (0.7)	.46
I found the using (SI or VT) visualization to be practical for use.	2.4 (0.9)	2.8 (0.8)	2.6 (0.8)	.09
I could imagine performing interventions with AR support using (SI or VT) visualization.	2.8 (1.1)	2.7 (0.9)	2.7 (1.0)	.57
I believe that AR[c] support through (SI or VT) visualization enhances patient safety.	2.5 (0.9)	2.7 (0.9)	2.6 (0.9)	.19
I was generally satisfied with the AR support through the using (SI or VT) visualization.	2.8 (0.9)	2.9 (0.8)	2.8 (0.8)	.66

[a]SI: superimposition.

[b]VT: virtual twin.

[c]AR: augmented reality.

Table . Summarized open questions.

Visualization modalities	Positive	Negative
Superimposition	• 3D, intuitive[a,b] • A novelty experience[a] • Accurate[a,b] • Beginner-friendly[a,b] • Clear and detailed[a] • Contrasting colors enhance structural differentiation[b] • Could be observed in all directions[b] • Easy localization of structures[a,b] • Easy to use[a,b] • Feeling of safety[b] • Free of time delay[a] • Good guidance and spatial relationship[a,b] • Inner structures could be easily seen in all directions[a] • Potential to simplify the process[a] • Simple design[a] • Time-saving[a,b]	• Depth is perceived differently in different angles[a,b] • Difficult to map 3D structures to 3D surface[a,b] • Hard to identify the position of the structures[a] • Hard to recognize the tip of the pen and place to draw[a,b] • Have to lock the registration and move the head phantom to fine-tune it to the hologram[a] • Inaccurate overlay, holograms are partially overlaid to the physical head phantom[a,b] • Need familiarization time[a] • Need to move the head phantom to overlay[b] • Not practical[a] • Relatively lacks sharpness[a] • Restriction of viewpoint, cannot rotate the head phantom to observe after locking the registration[a] • Some structures have merged[a,b] • The guidance makes the user neglect the critical anatomical landmarks, causing imprecise localization[a] • The hologram is blurred, and the double image is tiring[a] • The position of the head phantom and the participant should be kept constant[a]

Visualization modalities	Positive	Negative
Virtual twin	• 3D visualization, intuitive[a] • Accurate[a] • Assistive setup[a,b] • Clear visualization of the anatomical structures' location[a,b] • Direct views[a] • Easy to use[a,b] • Good guidance[a,b] • Guidance to the targets depends on the distance to the landmarks[a] • Head could be moved to draw[a] • Intuitive[a] • Less irritating than SI[b] • Like working with a textbook on the side[a] • No registration problem, 3D model hardly disturbs as an aid[a] • Only exit points are good to paint and recognizable[a] • Opportunity to apply to other structures[b] • With improved guidance, anatomical structures can be localized more effectively, referring to reference only when necessary[a,b] • Without overlapping, both the pen and the drawing position are clearly visible[a,b]	• Better able to rotate or zoom in[a] • Better to reposition or move the model without moving yourself[a] • Confusing and inaccurate[a] • Deficiency of necessary landmarks[b] • Hard to estimate where to draw[a] • Image lacking sharpness[a] • Lack of 3D guidance[a] • Limited transferability to the real head[b] • Little added value compared to drawing according to anatomical landmarks[a] • Localization cannot be tracked as precisely as with SI[b] • Longer time for eyes to adapt to[a] • Need to turn around the head phantom and hard to find the correct position[a] • Not practical for clinical use[a] • Possible spatial discrepancy, inaccurate drawing, impractical[a,b] • Required more cognitive effort compared to direct projection[b] • Rotation of the virtual head is restricted[a] • Slower in time[a,b] • Spatial depth is hard to estimate[b] • Switching attention back and forth between the head and hologram is confusing[a] • The smooth white head phantom offers few points of reference, hard to transfer the anatomical structures[b]

[a]Dental and medical student.

[b]Oral and maxillofacial, plastic, and oral surgeons.

Discussion

Principal Findings

We systematically evaluated the localization accuracy between two visualization modalities: SI with markerless inside-out tracking and VT for different types of anatomical structures in the head and neck region. The primary endpoint (absolute accuracy of 0D structure) revealed that SI was significantly more accurate than VT by 1.4 mm (P=.003). In terms of relative accuracy of 0D point structures, SI also outperformed VT by a margin of 2.6 mm (P<.001). VT showed comparable accuracy for 2D structures and notably superior accuracy (ASD, P=.01; HD, P=.03) for 3D structures, although it required an additional 16 seconds on average (P=.02). Likert questions revealed comparable results between two modalities. Feedback from open-ended questions (Table 3) highlighted SI for ease of understanding, intuitiveness, and time efficiency, yet noted persistent challenges with depth perception, visual occlusion, and virtual-real misalignment. Conversely, VT was perceived as simpler, clearer, and free of occlusion and misalignment issues, despite lacking direct positional cues on real head phantoms and requiring frequent attention shifts between physical and virtual models. Overall, user preferences were evenly split, reflecting comparable experiences despite each modality's distinct strengths and limitations.

Respective Strengths and Weaknesses

In contrast to VT without tracking, the accuracy of SI depends on the inside-out tracking of the HMD used and can be attributed to 3 main factors, namely the registration accuracy of the tracking (Vuforia), the spatial mapping performance (HL2), and the visual occlusion [27]. Previous studies illustrated Vuforia software development kit's registration in the HL2 highly depended on the richness of the shape and texture of the tracked target and ranged from less than 2 mm to more than 10 mm for translational error [19,28], which can propagate into an angular deviation of the task-specific cutting plane up to 14.7° [20]. Furthermore, Vuforia tracking is sensitive to environmental light intensity, distance to the target object, and the extent of the surface covered [29]. In addition, HL2 used visual inertial-simultaneous localization and mapping (VI-SLAM) to continuously map the environment and update its position and orientation within a global coordinate system, anchoring virtual content to real-world features [30,31]. However, VI-SLAM's accuracy can be affected by factors, such as pose prediction latency, user motion, environment, and sensor fusion, such as poor integration between the red, green, blue camera and inertial measurement unit [30,31]. This VI-SLAM error accumulated along the way, reaching 5 mm per 628 mm traveled in the clinical environment [31,32]. Moreover, the jitter latency caused by such sensor fusion could further compromise user experience, increase cognitive load, and induce fatigue [30,33]. Last but not least, visual occlusion, where virtual objects can obstruct or distort the view of the physical counterpart, further compromises the accuracy of SI. Many participants reported difficulty in identifying the position of the pen, drawn line, and occluded virtual content, which was also observed in another study [34]. This occlusion problem could lead to severe damage during surgery by overlooking anatomical structures and events [35,36]. All these factors together may contribute to the overall accuracy achieved by SI.

On the other hand, VT showed comparable accuracy (inferior in 0D, comparable in 2D, and superior in 3D structures) to SI with markerless inside-out tracking of the HMD, but without the aforementioned problems of SI. This was largely due to VT's design, which bypassed the need for precise virtual-real overlay or accurate anchoring by displaying the virtual model next to the real head phantom. Nevertheless, VT as a visualization modality free of misalignment, unaffected by occlusion, and less sensitive to spatial mapping instability could substitute SI in macro localization tasks. Since VT lacked direct positional cues to guide localization, it likely depended on the surgeon's ability to estimate distance, where surgeons performed an average error of 1.4 (SD 1.2) mm in 5 mm and 2.0 (SD 1.9) mm in 1 cm estimation in a research [37]. The distances between the nerve exit points in our study were approximately 4 cm between supraorbital and infraorbital foramina and 7 cm between infraorbital and mental foramina. If we assume the estimation error was in a linear model, this corresponded to a mean error of 5.6-9.2 mm, which aligned with VT's average relative accuracy (mean 6.0, SD 5.0 mm), inferior to SI (mean 3.4, SD 2.2) mm). Therefore, one could argue that SI is only meaningful if its accuracy exceeds the limits of human distance estimation.

Comparison to Prior Work

In scenarios where precise localization is required, such as orbital fracture reconstruction or trajectory drilling, optical tracking remains the most accurate method to date [15]. Consequently, numerous studies have adopted optical tracking to optimize registration of SI. For instance, Tu et al [38] achieved entry point accuracy of mean 2.8 (SD 1.3) mm and angular accuracy of mean 3.0° (SD 1.2°), optimizing registration accuracy to mean 2.0 (SD 0.7) mm through optical tracking. Similarly, Iqbal et al [39] combined the HL2 built-in camera with an external optical tracking camera, further reducing translation and rotation errors to 2.1 mm and 1.5°, respectively. In contrast, VT with external optical tracking could also visualize both the virtual instrument and the target anatomy but adjacent to the patient in real-time. This framework achieved higher accuracy than the aforementioned SI systems and comparable accuracy to SNS, with translational deviations of mean 0.9 (SD 0.4) mm and mean 1.0 (SD 0.5) mm at entry and end points, respectively, and a rotational deviation of mean 1.1° (SD 0.6°) [15], within the clinically feasible range (~2 mm) [12]. The noticeable difference between VT by 0.9 (SD 0.4) mm and SI by 2.1 mm with a similar optical tracking framework likely resulted from the aforementioned factors, such as registration errors, VI-SLAM instability, jitter, and visual occlusion. This raises the question of whether SI with optical tracking should be considered the optimal AR visualization modality for surgical scenarios, particularly given that VT achieved similar accuracy under similar tracking conditions without encountering these limitations.

However, all these values assume that localization accuracy is measured in anatomically exposed structures, where perfect localization could theoretically reach 0 mm. However, this ignores a crucial aspect of real-world scenarios: anatomical structures are typically covered by tissue, which prevents direct access and inherently limits localization accuracy. In our study, this was particularly relevant due to the soft and bone tissue overlying the anatomical target structures (ie, nerve exit points, inferior alveolar nerves, and salivary glands). According to the literature, the average soft tissue thickness of the head, face, and neck is 9.4 (SD 6.2; range 2.4-28.1 mm) mm in women and 10.5 (SD 7.2; range 2.7-32.4) mm in men [40]. Thus, the measured localization accuracies observed with SI and VT for nerve exit points cannot be directly compared to classical navigation scenarios with fully exposed anatomical targets, as they are composed of 4 main influencing factors

Localization Accuracyij=β1·Overlaying Tissue Thicknessi Anatomical Constraint+

β2·Modalityij Visualization Modality (SI vs. VT)+uj Subject Bias+εij Residual Error (AR Noise)

First, the tissue thickness overlaid the target structure. This resulted in a decrease in localization accuracy of 1.4 mm per 1 mm of overlying tissue thickness (LMM, unstandardized coefficient β_1=1.4; *P*<.001). Then there is the influence of the VT modality, which added an additional error of 1.4 mm compared to SI (LMM, β_2=1.4; *P*=.003), and the average participant-specific bias, which was 0.8 mm (average magnitude of random intercept for individuals in the LMM). Finally, the residuals described the general pattern of localization error, with a mean absolute error of 1.8 (SD 2.9) mm for SI and 2.5 (SD 3.5) mm for VT.

For line- or volume-based structures, this correlation could not be reliably captured by the model. This was probably because localization accuracy depended not only on the viewing direction but also on dynamic changes in the perceived target margin along that direction. This contrasts with the single-point structure, which is invariably depicted as a point in all directions. As a result, the localization accuracy for line- and volume-based structures is biased by viewing direction and margin variability, in addition to the tissue thickness. As in Van Gestel et al [41], where a brain tumor was dynamically projected onto the skin along a vector from its center to the instrument tip, the participant's line of sight in our study played a comparable role to the instrument tip. As the viewing angle shifted, the visible margin of the gland changed in real time, introducing variability in the drawn curves and affecting both ASD and HD. Even for targets on the skin, like in wound area estimation using photography, variation in camera angle could introduce 10% error [42]. Although we could not directly quantify viewing angles and changing margins, aligning the participant's gaze with the vector from the structure's centroid to its nearest skin projection may help minimize delineation errors related to such bias.

Clinical Implications

First, VT appears particularly advantageous for tasks requiring coarse localization and stable spatial orientation. VT provides a reliable anatomical context and could help mitigate cognitive errors, such as confusion of lateral sides or anatomical levels. These errors often arise in apparent symmetrical regions, especially in the absence of clear preoperative marking or adequate visual guidance. For example, in thoracolumbar spine surgery, reliance solely on intraoperative fluoroscopy may be insufficient to reliably distinguish vertebral levels, especially in the presence of anatomic variants, inadequate intraoperative imaging fields, and unreliable surface landmarks, with 50% - 67% of surgeons reporting such errors [43,44]. VT could orient the surgeon by allowing the user to align CT-based virtual models with the patient's posture, enabling clear visualization of the spine and reducing wrong-level or wrong-side misorientation. Second, in maxillofacial reconstruction, the VT technique offers significant value by displaying planned bone segments and prebent fixation plates alongside the operative field. This side-by-side visualization enables real-time comparison and intraoperative adjustment of plate bending, reducing the need for repeated fitting at the surgical site as standard techniques do [45], and thereby lowering the risk of infection. Compared with preoperative 3D printing, such a technique could also minimize fabrication time and offer greater flexibility for intraoperative adjustments. In the following free flap reconstruction procedures, VT offers robustness in environments prone to bleeding, swelling, or tissue deformation, where SI overlays can drift or become unreliable. By anchoring the virtual model generated from virtual surgical planning adjacent to the surgical site, VT provides a stable frame of reference with consistent skeletal landmarks, even when soft tissues shift [46]. Third, VT is well-suited to fractures and postoncologic defects of the orbit and midface requiring symmetry (eg, zygomatic arch, orbital floor, and medial wall) [47]. By rendering the contralateral mirrored anatomy, target orbital volume, planned implant contour, and craniofacial buttresses adjacent to the field, the surgeon could continuously compare the intraoperative reduction with the surgical plan.

While VT may help reduce orientation errors, SI demonstrates its strength in scenarios that demand high-precision localization. For example, in mandibular reconstruction surgery using the anterolateral femoral flap, accurate localization of the perforator vessels is crucial to flap viability and surgical success. One study found the SI with remote-controlled overlay (mean 3.5, SD 2.8 mm) achieved significant superior localization accuracy in anterolateral femoral perforator vessels than ultrasonic color Doppler (mean 9.6, SD 5.8 mm; P<.001) [48]. Our findings showed that SI had clear advantages in point-based localization tasks. This feature is particularly important in procedures such as sentinel lymph node biopsy, where accurately identifying nodes just a few millimeters beneath the surface is crucial for surgical success. Duan et al [49] reported that AR SI with motion compensation achieved sub-3 mm localization error in melanoma sentinel lymph node biopsy. Moreover, SI demonstrated significantly superior relative accuracy (P<.001). This is because, despite the offset, SI preserved the spatial relationships between landmarks. The scenario that benefits from this strength is when relative distances between anatomical points must be accurately estimated, especially when a landmark has already been explored and exposed. For example, in head and neck tumor surgeries, surgeons often use the tragal pointer as a surgical landmark to identify the facial nerve trunk and the maxillary artery during procedures, such as parotidectomy, mandibular osteotomy, and temporomandibular joint arthroplasty [50]. In addition, in skull base surgery, surgeons often rely on stable bony landmarks, such as the occipital condyle or mastoid process, to sequentially locate cranial nerve exit points, including the jugular foramen and hypoglossal canal [51].

SI with markerless inside-out tracking and VT could be combined across different stages of the tasks. First, VT provides general spatial awareness, such as adapting to specific patient positioning, orienting with comprehensive medical imaging, or selecting approximate entry points. Once a key anatomical landmark is exposed, SI could rapidly guide surgeons to adjacent structures by using relative spatial relationships, minimizing the need for repeated attention switching [52,53]. If SI causes visual obstruction, cognitive overload, or registration instability and inaccuracy, SI can be temporarily deactivated, allowing VT to take over as a stable spatial reference. This hybrid modality enables adaptive assistance, providing surgeons with tailored support at different procedure stages based on clinical needs.

Our findings showed that user preferences were almost evenly split between SI and VT, underscoring the limitations of relying on either visualization method in isolation. Rather than competing alternatives, SI and VT could be viewed as complementary tools that respond to different scenarios. While SI enables precise overlay of subcutaneous landmarks, VT provides more reliable orientation under deformation or registration drift. These complementary features suggest that future AR systems should integrate both approaches within a single workflow.

Limitations

This study has some limitations. First, the polystyrene foam head phantoms used in the experiment lacked realistic features, such as skin texture, natural color, and anatomical details, which are critical for accurate identification of anatomical landmarks in the real clinical scenarios. However, using these phantoms allowed for reproducible evaluation of the performance of two modalities. Second, the homogeneous and rigid phantom surface may have favored SI by registration. Unlike real surgical environments, phantoms lack deformable soft tissues, surgical draping, fluids, and light reflections, all of which can substantially increase registration and tracking errors for SI [54,55]. In contrast, VT does not require accurate overlay; thus, it was not hindered by those problems. These considerations suggest that the relative advantage of SI observed in phantom experiments may be attenuated in vivo, whereas VT could perform more robustly in real surgical settings. Third, in real clinical scenarios, the phantom's components, such as the mandible, could not replicate the mobility of human anatomy. This mobility may pose a significant challenge to markerless inside-out registration and further accurate anatomical localization for SI. In contrast, mobile parts in VT may be a potential solution. To address these challenges, cadaver studies or studies with high-fidelity phantoms replicating the mobility of anatomical structures should be conducted to validate the clinical applicability and generalizability of the findings. Fourth, since the difference in absolute accuracy in the sample size calculation was less than 5 mm, the study may have been underpowered to detect the influence of some fixed effects. Subsequent studies should consider increasing the sample size to enhance statistical power and generalizability. Finally, current findings are constrained to the facial region, where underlying bone structures provide a stable spatial reference. It would be valuable to investigate the performance of two modalities in other regions of the body like the abdomen, where soft tissue may bring additional challenges.

Future Directions

In addition to further validation with cadaveric studies or high-fidelity phantoms, future work should also address technical factors that directly influence localization accuracy. In particular, subcutaneous soft tissue thickness, variations in viewing perspective, and the resulting margin variability were shown to pose consistent challenges for both AR modalities. To mitigate these effects, new visualization approaches need to be developed to reduce the effects of viewing perspective and account for the effects of the overlying tissue, regardless of the visualization modality. First, the user's viewing angle could be guided in AR. One possible strategy would be to create a virtual cylindrical tunnel of 2 circles between the target structure and the skin surface, orienting the user to view in a planned direction. Second, the AR visualization should establish a clear connection between the overlying tissue and the target structures, for example, for nerve exit points, a line connecting the points and their planned skin projection, clearly identifying the planned margin and mitigating the inaccuracy introduced by the overlying tissue.

While these approaches address specific visualization challenges, the next step lies in advancing toward a hybrid, context-aware AR system. With advances in registration accuracy, hardware performance, and integration of AI technologies, such a system could autonomously detect procedural phases, surgical context, and anatomical exposure. Based on this contextual understanding, it could dynamically switch between VT and SI modes, providing global spatial orientation and reference by VT and precise overlays for local structure localization by SI. This intelligent modality would reduce cognitive load and enable phase-specific surgical guidance.

Conclusion

This study systematically compared SI with markerless inside-out tracking and VT for surgical localization tasks in the head and neck region. SI demonstrated superior localization accuracy in 0D structures, whereas VT revealed robust spatial orientation, comparable accuracy in 2D, and superior accuracy in 3D structures. These complementary strengths suggested that VT represents a viable alternative for macro localization, while SI may be preferable for fine-grained, sequential landmark tasks. Rather than assuming SI to be universally applicable across all surgical contexts, our findings emphasize the need for context-adaptive AR strategies that can dynamically leverage the strengths of both modalities.

Acknowledgments

We sincerely thank all participating students and surgeons for their valuable contributions to this study. Open access funding provided by the Open Access Publishing Fund of RWTH (Rheinisch-Westfälische Technische Hochschule) Aachen University.

Funding

This work was supported by the REACT-EU project KITE (grant number EFRE-0801977, Plattform für KI-Translation Essen (https://kite.ikim.nrw/), FWF enFaced 2.0 (grant number KLI-1044 (https://enfaced2.ikim.nrw/), Clinician Scientist Program of the Faculty of Medicine RWTH Aachen University (BHP). CG was funded by the Advanced Research Opportunities Program (AROP) of RWTH (Rheinisch-Westfälische Technische Hochschule) Aachen University.

Data Availability

The datasets generated or analyzed during this study, as well as the Python scripts used for data processing and analysis (stored on GitHub [58]), are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: BHP, CG
Methodology: CG, YL
Software: CG, GL, YL
Validation: YL, KG, KX
Formal analysis: YL, BHP, KX
Investigation: YL, KG, GL, KX
Resources: BHP, JE, GL, FH, RR

Data curation: YL, KG, GL, KX
Writing–original draft: YL
Writing–review and editing: BHP, YL, KX, GL, CG, KG, AB, FH, RR, MdlF, JE
Visualization: YL, BHP
Supervision: BHP, JE
Project administration: BHP
Funding acquisition: CG, BHP

Conflicts of Interest

BHP is an associate editor of the *Journal of Medical Internet Research*. All other authors declare no other conflicts of interest.

Checklist 1
CONSORT checklist.
[PDF File, 68 KB - games_v14i1e75962_app1.pdf]

References

1. Christison-Lagay E. Complications in head and neck surgery. Semin Pediatr Surg 2016 Dec;25(6):338-346. [doi: 10.1053/j.sempedsurg.2016.10.007] [Medline: 27989359]
2. Nadershah M, Salama A. Removal of parotid, submandibular, and sublingual glands. Oral Maxillofac Surg Clin North Am 2012 May;24(2):295-305. [doi: 10.1016/j.coms.2012.01.005] [Medline: 22364886]
3. Bachar G, Siewerdsen JH, Daly MJ, Jaffray DA, Irish JC. Image quality and localization accuracy in C-arm tomosynthesis-guided head and neck surgery. Med Phys 2007 Dec;34(12):4664-4677. [doi: 10.1118/1.2799492] [Medline: 18196794]
4. Wippold FJ. Head and neck imaging: the role of CT and MRI. J Magn Reson Imaging 2007 Mar;25(3):453-465. [doi: 10.1002/jmri.20838] [Medline: 17279529]
5. Ishibashi M, Fujii S, Kawamoto K, et al. The ability to identify the intraparotid facial nerve for locating parotid gland lesions in comparison to other indirect landmark methods: evaluation by 3.0 T MR imaging with surface coils. Neuroradiology 2010 Nov;52(11):1037-1045. [doi: 10.1007/s00234-010-0718-1] [Medline: 20505928]
6. Glas HH, Kraeima J, van Ooijen PMA, Spijkervet FKL, Yu L, Witjes MJH. Augmented reality visualization for image-guided surgery: a validation study using a three-dimensional printed phantom. J Oral Maxillofac Surg 2021 Sep;79(9):1943. [doi: 10.1016/j.joms.2021.04.001] [Medline: 34033801]
7. Remschmidt B, Rieder M, Gsaxner C, Gaessler J, Payer M, Wallner J. Augmented reality–guided apicoectomy based on maxillofacial CBCT scans. Diagnostics (Basel) 2023 Sep 25;13(19):3037. [doi: 10.3390/diagnostics13193037] [Medline: 37835780]
8. Landaeta-Quinones CG, Hernandez N, Zarroug NK. Computer-assisted surgery: applications in dentistry and oral and maxillofacial surgery. Dent Clin North Am 2018 Jul;62(3):403-420. [doi: 10.1016/j.cden.2018.03.009] [Medline: 29903558]
9. Karnatz N, Möllmann HL, Wilkat M, Parviz A, Rana M. Advances and innovations in ablative head and neck oncologic surgery using mixed reality technologies in personalized medicine. J Clin Med 2022 Aug 16;11(16):4767. [doi: 10.3390/jcm11164767] [Medline: 36013006]
10. Gsaxner C, Pepe A, Wallner J, Schmalstieg D, Egger J. Markerless image-to-face registration for untethered augmented reality in head and neck surgery. : Springer; 2019 Presented at: Medical Image Computing and Computer Assisted Intervention (MICCAI 2019); Oct 13-17, 2019; Cham, Switzerland p. 236-244. [doi: 10.1007/978-3-030-32254-0]
11. Pepe A, Trotta GF, Mohr-Ziak P, et al. A marker-less registration approach for mixed reality–aided maxillofacial surgery: a pilot evaluation. J Digit Imaging 2019 Dec;32(6):1008-1018. [doi: 10.1007/s10278-019-00272-6] [Medline: 31485953]
12. Andrews CM, Henry AB, Soriano IM, Southworth MK, Silva JR. Registration techniques for clinical applications of three-dimensional augmented reality devices. IEEE J Transl Eng Health Med 2021;9:4900214. [doi: 10.1109/JTEHM.2020.3045642] [Medline: 33489483]
13. Pérez-Pachón L, Poyade M, Lowe T, Gröning F. Image overlay surgery based on augmented reality: a systematic review. Adv Exp Med Biol 2020;1260:175-195. [doi: 10.1007/978-3-030-47483-6_10] [Medline: 33211313]
14. Rose AS, Kim H, Fuchs H, Frahm JM. Development of augmented-reality applications in otolaryngology-head and neck surgery. Laryngoscope 2019 Oct;129(S3):S1-S11. [doi: 10.1002/lary.28098] [Medline: 31260127]
15. Li Y, Drobinsky S, Becker P, et al. Accuracy and efficiency of drilling trajectories with augmented reality versus conventional navigation randomized crossover trial. NPJ Digit Med 2024 Nov 10;7(1):316. [doi: 10.1038/s41746-024-01314-2] [Medline: 39523443]
16. Wolf J, Luchmann D, Lohmeyer Q, Farshad M, Fürnstahl P, Meboldt M. How different augmented reality visualizations for drilling affect trajectory deviation, visual attention, and user experience. Int J Comput Assist Radiol Surg 2023 Aug;18(8):1363-1371. [doi: 10.1007/s11548-022-02819-5] [Medline: 36808552]

17. Grossberg A, Mohamed A, Elhalawani H, et al. HNSCC. The Cancer Imaging Archive. 2020. URL: https://stage.cancerimagingarchive.net/collection/hnscc/ [accessed 2026-02-02]
18. Thabit A, Benmahdjoub M, van Veelen MLC, Niessen WJ, Wolvius EB, van Walsum T. Augmented reality navigation for minimally invasive craniosynostosis surgery: a phantom study. Int J CARS 2022 Aug;17(8):1453-1460. [doi: 10.1007/s11548-022-02634-y] [Medline: 35507209]
19. Frantz T, Jansen B, Duerinck J, Vandemeulebroucke J. Augmenting Microsoft's HoloLens with vuforia tracking for neuronavigation. Healthc Technol Lett 2018 Oct;5(5):221-225. [doi: 10.1049/htl.2018.5079] [Medline: 30464854]
20. Chui ECS, Mak KKK, Ng RHT, et al. Application of image recognition-based tracker-less augmented reality navigation system in a series of sawbone trials. Arthroplasty 2024 Aug 2;6(1):39. [doi: 10.1186/s42836-024-00263-1] [Medline: 39090719]
21. Vorobiov V. On 3D pose estimation for XR. classic approaches vs neural networks [Bachelor's thesis]. : Ukrainian Catholic University; 2021 URL: https://er.ucu.edu.ua/items/3696ed7e-49d4-45c4-bba4-be06fd4320eb [accessed 2025-12-03]
22. Palla B, Callahan N. Does the use of computer-assisted surgery affect the margin status in resections of ameloblastoma? J Oral Maxillofac Surg 2021 Jul;79(7):1467-1473. [doi: 10.1016/j.joms.2020.09.007] [Medline: 33031772]
23. Fowler J, Campanile Y, Warner A, et al. Surgical margins of the oral cavity: is 5 mm really necessary? J Otolaryngol Head Neck Surg 2022 Oct 4;51(1):38. [doi: 10.1186/s40463-022-00584-8] [Medline: 36195903]
24. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. J Stat Soft 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]
25. Dwan K, Li T, Altman DG, Elbourne D. CONSORT 2010 statement: extension to randomised crossover trials. BMJ 2019 Jul 31;366:l4378. [doi: 10.1136/bmj.l4378] [Medline: 31366597]
26. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332. [doi: 10.1136/bmj.c332] [Medline: 20332509]
27. Doughty M, Ghugre NR, Wright GA. Augmenting performance: a systematic review of optical see-through head-mounted displays in surgery. J Imaging 2022 Jul 20;8(7):203. [doi: 10.3390/jimaging8070203] [Medline: 35877647]
28. Allen DR, Peters TM, Chen ECS. A co-calibration framework for the accuracy assessment of vision-based tracking systems. In: Linte CA, Siewerdsen JH, editors. Presented at: SPIE Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling; Feb 20-24, 2022; San Diego, United States p. 598-608. [doi: 10.1117/12.2606815]
29. Sendari S, Firmansah A. Performance analysis of augmented reality based on Vuforia using 3D marker detection. Presented at: 2020 4th International Conference on Vocational Education and Training (ICOVET); Sep 19, 2020; Malang, Indonesia p. 294-298 URL: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=9229480 [accessed 2025-12-03] [doi: 10.1109/ICOVET50258.2020.9230276]
30. Wilmott JP, Erkelens IM, Murdison TS, Rio KW. Perceptibility of jitter in augmented reality head-mounted displays. Presented at: 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR); Oct 17-21, 2022; Singapore, Singapore p. 470-478 URL: https://ieeexplore.ieee.org/document/9995228 [accessed 2025-12-03] [doi: 10.1109/ISMAR55827.2022.00063]
31. Feigl T, Porada A, Steiner S, Löffler C, Mutschler C, Philippsen M. Localization limitations of ARCore, ARKit, and Hololens in dynamic large-scale industry environments. Presented at: 15th International Conference on Computer Graphics Theory and Applications; Feb 27-29, 2020; Valletta, Malta p. 307-318. [doi: 10.5220/0008989903070318]
32. Matyash I, Kutzner R, Neumuth T, Rockstroh M. Accuracy measurement of HoloLens2 IMUs in medical environments. Curr Dir Biomed Eng 2021 Oct 1;7(2):633-636. [doi: 10.1515/cdbme-2021-2161]
33. Gsaxner C, Li J, Pepe A, Schmalstieg D, Egger J. Inside-out instrument tracking for surgical navigation in augmented reality. Presented at: VRST '21; Dec 8-10, 2021; Osaka, Japan p. 1-11 URL: https://dl.acm.org/doi/proceedings/10.1145/3489849 [accessed 2025-12-03] [doi: 10.1145/3489849.3489863]
34. Koyachi M, Sugahara K, Tachizawa K, et al. Mixed-reality and computer-aided design/computer-aided manufacturing technology for mandibular reconstruction: a case description. Quant Imaging Med Surg 2023 Jun 1;13(6):4050-4056. [doi: 10.21037/qims-22-1118] [Medline: 37284085]
35. Dixon BJ, Daly MJ, Chan H, Vescan AD, Witterick IJ, Irish JC. Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. Surg Endosc 2013 Feb;27(2):454-461. [doi: 10.1007/s00464-012-2457-3] [Medline: 22833264]
36. Dixon BJ, Daly MJ, Chan HHL, Vescan A, Witterick IJ, Irish JC. Inattentional blindness increased with augmented reality surgical navigation. Am J Rhinol Allergy 2014;28(5):433-437. [doi: 10.2500/ajra.2014.28.4067] [Medline: 25198032]
37. Conway RG, O'Neill N, Brown J, Kavic S. An educated guess - distance estimation by surgeons. Surg Open Sci 2020 Jul;2(3):113-116. [doi: 10.1016/j.sopen.2020.04.001] [Medline: 33981983]
38. Tu P, Wang H, Joskowicz L, Chen X. A multi-view interactive virtual-physical registration method for mixed reality based surgical navigation in pelvic and acetabular fracture fixation. Int J Comput Assist Radiol Surg 2023 Sep;18(9):1715-1724. [doi: 10.1007/s11548-023-02884-4] [Medline: 37031310]
39. Iqbal H, Rodriguez y Baena F. Semi-automatic infrared calibration for augmented reality systems in surgery. Presented at: Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); Oct 23-27, 2022; Kyoto, Japan p. 4957-4964. [doi: 10.1109/IROS47612.2022.9982215]

40. Yang W, He RK, Chen T, Haining W, Goossens R, Huysmans T. Soft tissue thickness estimation for head, face, and neck from CT data for product design purposes. In: Rebelo F, editor. Ergonomics In Design: AHFE International; 2022:558-565. [doi: 10.54941/ahfe1001983]
41. Van Gestel F, Frantz T, Buyck F, et al. Neuro-oncological augmented reality planning for intracranial tumor resection. Front Neurol 2023;14:1104571. [doi: 10.3389/fneur.2023.1104571] [Medline: 36998774]
42. Chang AC, Dearman B, Greenwood JE. A comparison of wound area measurement techniques: visitrak versus photography. Eplasty 2011 Apr 18;11:e18. [Medline: 21559060]
43. Epstein N. A perspective on wrong level, wrong side, and wrong site spine surgery. Surg Neurol Int 2021;12:286. [doi: 10.25259/SNI_402_2021] [Medline: 34221617]
44. Shah M, Halalmeh DR, Sandio A, Tubbs RS, Moisi MD. Anatomical variations that can lead to spine surgery at the wrong level: part III lumbosacral. Cureus 2020 Jul 28;12(7):e9433. [doi: 10.7759/cureus.9433] [Medline: 32864257]
45. Wilde F, Winter K, Kletsch K, Lorenz K, Schramm A. Mandible reconstruction using patient-specific pre-bent reconstruction plates: comparison of standard and transfer key methods. Int J Comput Assist Radiol Surg 2015 Feb;10(2):129-140. [doi: 10.1007/s11548-014-1065-1] [Medline: 24810111]
46. Metzler P, Geiger EJ, Alcon A, Ma X, Steinbacher DM. Three-dimensional virtual surgery accuracy for free fibula mandibular reconstruction: planned versus actual results. J Oral Maxillofac Surg 2014 Dec;72(12):2601-2612. [doi: 10.1016/j.joms.2014.07.024] [Medline: 25315311]
47. Khaqani MS, Tavosi F, Gholami M, Eftekharian HR, Khojastepour L. Analysis of facial symmetry after zygomatic bone fracture management. J Oral Maxillofac Surg 2018 Mar;76(3):595-604. [doi: 10.1016/j.joms.2017.10.005] [Medline: 29121488]
48. Wei B, Lu G, Bai Z, et al. Augmented reality in preoperative anterolateral thigh flap perforators positioning: a pilot diagnostic study. Oral Oncol 2025 Mar;162:107189. [doi: 10.1016/j.oraloncology.2025.107189] [Medline: 39954606]
49. Duan H, Yang Y, Niu WL, et al. Localization of sentinel lymph nodes using augmented-reality system: a cadaveric feasibility study. Eur J Nucl Med Mol Imaging 2025 Aug;52(10):3643-3652. [doi: 10.1007/s00259-025-07216-z] [Medline: 40123008]
50. Muhleman MA, Wartmann CT, Hage R, et al. A review of the tragal pointer: anatomy and its importance as a landmark in surgical procedures. Folia Morphol (Warsz) 2012 May;71(2):59-64. [Medline: 22648581]
51. Thintharua P, Chentanez V. Morphological analysis and morphometry of the occipital condyle and its relationship to the foramen magnum, jugular foramen, and hypoglossal canal: implications for craniovertebral junction surgery. Anat Cell Biol 2023 Mar 31;56(1):61-68. [doi: 10.5115/acb.22.105] [Medline: 36635090]
52. Liebmann F, Roner S, von Atzigen M, et al. Pedicle screw navigation using surface digitization on the Microsoft HoloLens. Int J Comput Assist Radiol Surg 2019 Jul;14(7):1157-1165. [doi: 10.1007/s11548-019-01973-7] [Medline: 30993519]
53. Pietruski P, Majak M, Światek-Najwer E, et al. Supporting mandibular resection with intraoperative navigation utilizing augmented reality technology - a proof of concept study. J Craniomaxillofac Surg 2019 Jun;47(6):854-859. [doi: 10.1016/j.jcms.2019.03.004] [Medline: 30914226]
54. Schlueter-Brust K, Henckel J, Katinakis F, et al. Augmented-reality-assisted K-wire placement for glenoid component positioning in reversed shoulder arthroplasty: a proof-of-concept study. J Pers Med 2021 Aug 10;11(8):777. [doi: 10.3390/jpm11080777] [Medline: 34442421]
55. Suter D, Hodel S, Liebmann F, Fürnstahl P, Farshad M. Factors affecting augmented reality head-mounted device performance in real OR. Eur Spine J 2023 Oct;32(10):3425-3433. [doi: 10.1007/s00586-023-07826-x] [Medline: 37552327]
56. AR-VIZ [accessible by request]. GitHub. URL: https://github.com/OMFSdigital/AR-VIZ [accessed 2025-12-24]

Abbreviations

AR: augmented reality
ASD: average surface distance
CONSORT: Consolidated Standards of Reporting Trials
CT: computed tomography
HD: Hausdorff distance
HL2: HoloLens 2
LMM: linear mixed-effects model
NASA-TLX: National Aeronautics and Space Administration Task Load Index
RWTH: Rheinisch-Westfälische Technische Hochschule
SNS: surgical navigation systems
VI-SLAM: visual inertial-simultaneous localization and mapping
VT: virtual twin

Edited by A Coristine; submitted 22.Apr.2025; peer-reviewed by E Berglund, Y Yang; revised version received 21.Oct.2025; accepted 21.Oct.2025; published 13.Jan.2026.

Please cite as:
Li Y, Luijten G, Gsaxner C, Grunert K, Bader A, Hölzle F, Röhrig R, de la Fuente M, Egger J, Xie K, Hinrichs-Puladi B
Usability Study of Augmented Reality Visualization Modalities on Localization Accuracy in the Head and Neck: Randomized Crossover Trial
JMIR Serious Games 2026;14:e75962
URL: https://games.jmir.org/2026/1/e75962
doi:10.2196/75962

Co-Designing Mobile Serious Games to Support Patients With Psoriatic Arthritis and Chronic Pain: Mixed Methods Study

Bárbara Ramalho[1], MSc; Samuel Gomes[1], MSc; Marta Silva Vicente[1], MSc; Filipa Magalhães[1], MSc; Rodolfo Gonçalves Costa[1], MSc; Sandra Gama[2], PhD; Vasileios Charisis[3], PhD; Leontios Hadjileontiadis[3,4], PhD; Sofia B Dias[5], PhD; iPROLEPSIS Consortium[3]

[1]HUMAN Lab, INESC-ID, Instituto Superior Técnico, Faculdade de Motricidade Humana, Universidade de Lisboa, Lisbon, Oeiras, Portugal

[2]HUMAN Lab, INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

[3]Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Central Macedonia, Greece

[4]Department of Biomedical Engineering and Biotechnology, Khalifa University, Abu Dhabi, United Arab Emirates

[5]Center of Interdisciplinary Study of Human Perfomance (CIPER), Faculdade de Motricidade Humana, Universidade de Lisboa, Estrada da Costa, 1495-751 Cruz Quebrada, Lisbon, Portugal

Corresponding Author:
Sofia B Dias, PhD
Center of Interdisciplinary Study of Human Perfomance (CIPER), Faculdade de Motricidade Humana, Universidade de Lisboa, Estrada da Costa, 1495-751 Cruz Quebrada, Lisbon, Portugal

Abstract

Background: Serious games offer promising avenues for clinical care by enhancing patient engagement and delivering therapeutic benefits. In psoriatic arthritis (PsA), chronic pain contributes to emotional distress, functional limitations, and reduced well-being. While symptom-tracking apps exist, few digital interventions directly address chronic pain through engaging, therapeutic experiences tailored to patients' cognitive and physical needs.

Objective: This study aimed to co-design mobile serious games—NoPain Games—to support patients with PsA in managing chronic pain. Conducted within the iPROLEPSIS Horizon Europe project, the study involved a multidisciplinary cocreation session with rheumatologists, researchers, and technical experts, followed by a usability feedback session with patients with PsA. The goal was to identify therapeutic priorities, refine game mechanics, and assess usability to inform the development of personalized, accessible digital interventions.

Methods: A sequential mixed methods design was used. First, a 90-minute remote cocreation session was held with 14 experts (6 rheumatologists, 4 technical experts, and 4 researchers) from 3 European countries. Participants reviewed game storyboards and discussed therapeutic and design priorities. Thematic analysis of transcribed discussions identified key insights. Next, a 60-minute remote usability feedback session was conducted with 5 patients with PsA (aged 25 - 64 y), who interacted with 2 high-fidelity game prototypes (Space Oddity and Four Seasons). Usability was assessed using the System Usability Scale (SUS), and qualitative feedback was collected through moderated discussion. Item-level analysis using item characteristic curves provided deeper insight into usability perceptions and item sensitivity. All ethics requirements were met for both study phases.

Results: Thematic analysis of the transcribed dialogs of the cocreation session revealed three core themes: (1) therapeutic benefits (pain distraction, memory enhancement, cognitive stimulation, stress reduction, and creative engagement); (2) game difficulty (balancing duration and complexity to sustain engagement without fatigue); and (3) accessibility and interaction (addressing physical limitations, optimizing touchscreen usability, and ensuring inclusive design). These insights informed the development of 2 NoPain Game prototypes, which received a SUS score of 79 (SD 10.4; 95% CI 69.89 - 88.11), indicating good usability. Item characteristic curve analysis showed strong discrimination for learnability, while ease of use and confidence exhibited ceiling effects. Items like support needs and inconsistency showed minimal variability, and the learning curve demonstrated delayed but meaningful responsiveness at higher usability levels. Qualitative feedback reinforced the relevance of difficulty adjustment, technical refinements, and game mechanics, offering actionable insights for future iterations and broader implementation.

Conclusions: This study uniquely contributes to the field by co-designing mobile serious games specifically for patients with PsA, integrating expert and patient input to address chronic pain through accessible, cognitively engaging digital interventions—an approach not previously explored in this population. Future work will refine game mechanics, integrate adaptive difficulty, and conduct clinical trials to evaluate therapeutic efficacy. NoPain Games may serve as complementary tools within digital care ecosystems, offering support tailored to the therapeutic needs and physical limitations of patients with PsA underserved by conventional therapies.

(JMIR Serious Games 2026;14:e75072) doi:10.2196/75072

KEYWORDS
serious games; psoriatic arthritis; chronic pain; NoPain Games; mHealth; digital health care; iPROLEPSIS; mobile health

Introduction

In health care, serious games—designed for purposes beyond entertainment—have generated considerable interest and sparked ongoing debate [1]. Their potential has often been overlooked due to the still prevailing perception of games as purely recreational objects rather than valuable tools for skill development and health rehabilitation [2]. Research studies, however, have demonstrated the benefits of serious games, particularly their ability to reduce stress [3], enhance mobility [4,5], and alleviate pain [6,7].

Many of these benefits align with the symptomatology of rheumatic diseases, making it possible to use serious games as health care tools for conditions such as psoriatic arthritis (PsA). PsA, a degenerative rheumatic disease, is marked by joint stiffness and persistent pain, significantly affecting patients' quality of life [8]. As there is no cure, treatment primarily focuses on managing symptoms and holistically improving patients' well-being, often through interventions like physical therapy [9]. However, these traditional approaches can feel repetitive or burdensome for patients, leading to low adherence and reduced effectiveness. While various digital tools, particularly mobile apps for disease management, have been proposed [10], they primarily serve as symptom-tracking solutions and fail to address the ongoing joint pain experienced by patients.

Serious games offer a compelling alternative to complement traditional treatments by creating experiences that are both engaging and therapeutic [11]. For example, DaktylAct, a touch-based serious game, has been proposed as an innovative tool for assessing fine motor skills in PsA using novel digital biomarkers [12]. Furthermore, immersive experiences in chronic pain management have been extensively studied, with strong evidence supporting their pain-relieving capabilities, particularly in virtual reality (VR) environments [6,13]. Building on this, previous research has highlighted the effectiveness of exergames in altering pain perception by engaging multiple sensory modalities [14]. For instance, Gold et al [13] highlighted how VR can influence neurobiological mechanisms, showing that auditory, visual, and tactile stimuli reduce activity in pain-processing areas. In addition, Hoffman et al [15] further validated the effectiveness of immersive digital interventions by comparing VR and opioids in pain relief. Their findings suggest that while both approaches are beneficial independently, their combination leads to even greater analgesic effects. These findings, derived from subjective pain scales and functional magnetic resonance imaging, highlight the potential of mobile-based alternatives with similar engagement mechanisms to perform effectively without the need for specialized hardware.

Moreover, recent studies have further expanded the evidence base for serious games in pain management. Beltran-Alacreu et al [16] developed a task-oriented serious game for older adults with chronic neck pain, demonstrating its suitability and therapeutic potential. Saragih et al [7] conducted a systematic review and meta-analysis confirming the efficacy of serious games in managing chronic pain among older adults, reinforcing their clinical relevance. Additionally, Peña et al [17] explored the use of digital art and attachment priming in a web-based serious game, showing promising results in reducing both pain and social disconnection in individuals with chronic pain and loneliness.

Immersion remains a critical factor in pain management games. Gromala et al [6] emphasized that realistic visual environments, sound effects, and interactive storytelling contribute significantly to the overall therapeutic experience. Their concept of imaginary immersion introduces in-game threats and challenges to sustain player focus and engagement. We believe that translating these principles into mobile platforms requires innovative game mechanics that maintain immersion despite limited sensory channels. For instance, the MyRelief smartphone app presents a serious game to relieve chronic low-back pain through physical and psychological activities [18]. Although their analysis suggests that serious games can help improve patient status, the authors recognize that further research is needed to validate all game components.

Incorporating motivational elements in mobile pain management games is another critical design consideration. In this line, Ijaz et al [19] demonstrated the power of competitive motivation in their cycling game, which used on-screen scores of real and artificial competitors to enhance user engagement. Similarly, Tuah et al [20] reviewed common gamification elements in rehabilitation games, identifying key components such as points, leaderboards, badges, progression systems, and avatars that contribute to sustained participation.

To the best of our knowledge, no serious game has been specifically designed to target the chronic pain experienced by individuals with PsA. This study aims to explore the co-design and development of mobile serious games, referred to as NoPain Games, specifically tailored to support patients with PsA in managing chronic pain. Developed using an agile co-design methodology [21], these serious games represent a new approach to chronic pain management, as part of the Horizon Europe iPROLEPSIS project [22,23], which aims to create a personalized suite of games tailored to the needs of individuals with PsA. The primary objective of this study is to assess the therapeutic potential, usability, and design requirements of these games through a multidisciplinary cocreation process involving clinicians, researchers, and technical experts, along with the patients' usability feedback. By integrating cognitive stimulation, stress reduction, and inclusive interaction mechanics, the study seeks to establish foundational design principles for mobile-based digital interventions that enhance patient engagement and well-being. The underlying hypothesis is that co-designed serious games incorporating adaptive difficulty and accessible interfaces can serve as effective pain distractors and cognitive enhancers for individuals living with PsA.

Methods

Overview

In this study, we followed a sequential mixed methods design, integrating qualitative and quantitative data to inform the co-design and evaluation of mobile serious games for patients with PsA. This sequential design was selected to allow expert-derived qualitative insights to inform prototype development prior to patient usability testing. Integration occurred through iterative refinement of game mechanics based on themes from the cocreation session, followed by triangulation of usability scores and patient feedback to validate design decisions. In the qualitative phase, 14 experts (6 rheumatologists, 4 technical experts, 4 researchers) were recruited via purposive sampling from the iPROLEPSIS Consortium and participated in a 90-minute remote cocreation session. Eligibility required domain expertise in rheumatology, digital health, or game development. Participants reviewed game storyboards and contributed to design refinement through moderated discussion. In the quantitative phase, 5 patients with PsA (age 25 - 64 years) from 4 countries were recruited through targeted outreach. Eligibility required a confirmed PsA diagnosis and the ability to interact with mobile devices. Patients participated in a 60-minute remote usability session, interacting with 2 high-fidelity game prototypes. Usability was assessed using the System Usability Scale (SUS), and qualitative feedback was collected through a structured discussion. Study outcomes included: (1) identification of therapeutic priorities and design requirements (qualitative), (2) usability scores and item-level sensitivity analysis (quantitative), and (3) patient feedback on game experience and accessibility (qualitative). Thematic analysis followed Braun and Clarke’s framework, and SUS data were analyzed using descriptive statistics and item characteristic curves (ICCs). The study received ethical approval, and informed consent was secured from all participants.

The following subsections provide a detailed account of the study context, co-design procedures, participant characteristics, data collection instruments, ethics, and analytic strategies used to generate and interpret the findings.

Study Context

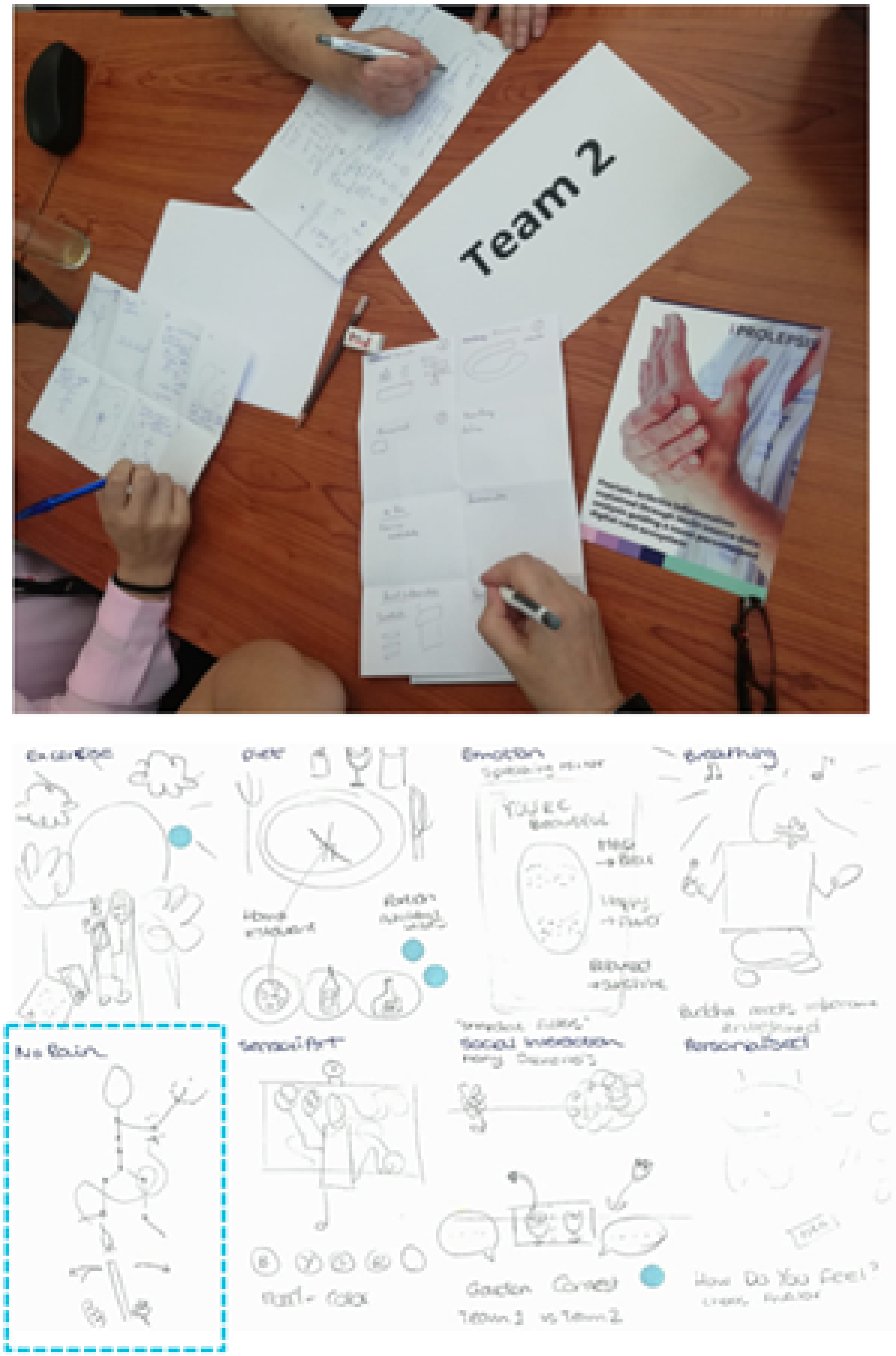
This study is part of the iPROLEPSIS Horizon Europe research project, which seeks to investigate the progression from general health to PsA through multi-source data analysis, ultimately creating an innovative, personalized digital care ecosystem. Within the framework of iPROLEPSIS, the project focuses on designing, developing, and validating cutting-edge digital biomarkers to assess and address PsA using an Artificial Intelligence-Personalized Game Suite (AI-PGS). Developed collaboratively with key stakeholders, the AI-PGS takes a comprehensive, multitargeted approach to managing PsA symptoms. It includes intervention activities aimed at enhancing breathing, mobility, stiffness, balance, coordination, fitness, diet, and mood. By embracing a holistic perspective, the AI-PGS addresses stress, anxiety, fatigue, and pain through various categories of serious games, such as NoPain Games, Exercise Games [24], Sensorimotor Art Games [25], Breathing Games [26], Dietary Games, and Emotional Games. This study emphasizes NoPain Games, showcasing the critical role of engaging patients in therapeutic activities that incorporate cognitive processes—such as memory, coordination, and visual perception tasks—to help alleviate and manage chronic pain and discomfort associated with PsA symptoms.

NoPain Games Agile Co-Design Process

The design and development process for the proposed games followed an agile approach [21], emphasizing early feedback to iteratively refine game concepts and prototypes. The ideation and storyboards phase will guide the foundation for subsequent stages, including prototyping, software development, and expert evaluation, while ensuring continuous cocreation efforts aimed at clinical validation (see Figure 1). Following the product backlog, sprint planning meetings, and backlog refinement phases, the design process began with the Crazy8s ideation approach [27]. More specifically, stakeholders—including patients, clinicians, researchers, and technicians—collaboratively sketched ideas for various game categories within the AI-PGS framework, including NoPain Games (see Figure 2). This phase captured diverse requirements and expectations, culminating in the initial development of a game design document outlining the NoPain Games’ key aspects, such as their concepts, mechanics, design, controls, and clinical relevance. This study highlights the subsequent agile cocreation sessions that brought together health care professionals, researchers, and technical experts as active cocreators and co-designers (see “Study participants and data collection” section) and patients with PsA as usability feedback providers (see “Feedback From Patients With PsA” section). In particular, using the storyboard technique [28], a collaborative session focused on refining and assessing the proposed designs for the NoPain Games (Figure 3), while another later session focused on the usability and general feedback of functional prototypes (Figure 4).

Figure 1. Schematic representation of the development framework for the proposed NoPain Games, illustrating the cocreation and agile methodologies used to achieve a minimum viable product. The process includes iterative 2 - 3 month sprints involving product backlog refinement (led by a technical coordinator), sprint planning, ideation, game development, expert evaluation, and clinical validation. All phases, except product backlog refinement, actively engaged patients, clinicians, researchers, and developers to ensure a multidisciplinary and user-centered approach.

Figure 2. Co-design session involving patients with psoriatic arthritis during the ideation process of the games using Crazy8s method. The initial sketch of the NoPain Games category is marked with a dashed blue square.



Initially Proposed NoPain Games Storyboards

Three different NoPain Games were initially designed: “Lightning,” “Flashing Stars,” and “The Garden” with related storyboards (Figure 3). In the “Lightning” game, players/patients are required to replicate a given pattern by lighting up rooms in a 2D house, reinforcing memory and coordination (Figure 3A). The “Flashing Stars” game engages players in counting falling stars within a night-sky scenario, fostering relaxation through gentle visual and cognitive stimulation (Figure 3B). Finally, “The Garden” game involves identifying a specific flower type within a vast field of blooms, promoting a sense of calm and concentration (Figure 3C). Overall, by incorporating the repetitive and rhythmic interactions of identifying flowers, lighting up rooms, or counting falling stars, the proposed NoPain Games encourage users to engage in predictable, structured movements that promote relaxation and a sense of control. These are simple and rewarding mechanics aimed to provide immediate feedback, reinforcing positive engagement, and promoting a sense of accomplishment, while enabling a calm environment. Designed for both iOS and Android operating systems, the proposed NoPain Games high-fidelity prototypes present similar interaction controls, with players/patients using a touchscreen for interaction, with calming audio to match each scenario.

Study Participants

Cocreation Session

To obtain feedback on the related storyboards, a cocreation session was held. The latter occurred in October 2023 and was conducted digitally via the Microsoft Teams platform with shared storyboard visuals. The session included 14 experts from 3 European countries, namely the United Kingdom, Greece, and Portugal. Participants were selected using purposive sampling [29], intentionally inviting individuals with expertise in relevant fields. They were identified through institutional affiliations within the iPROLEPSIS Consortium and invited via email by the coordinating team. Participation was based on voluntary engagement and without the provision of compensation. Exclusion criteria for experts included a lack of direct experience with PsA or digital health interventions. The group included 6 rheumatologists, 4 researchers, and 4 technical experts from the iPROLEPSIS Consortium, with multidisciplinary perspectives, enriching the discussions. The cocreation session was moderated by the last author, a qualitative researcher with expertise in participatory design. To minimize bias, the moderator did not participate in game development. Two qualitative coders (first and last author) independently reviewed transcripts and discussed potential biases during theme development (see Results).

Feedback Session

To maintain a patient-centered and clinically relevant design process, individuals living with PsA were actively engaged in a dedicated feedback session conducted in May 2025. During the session, participants evaluated a pilot version of Space Oddity and a design prototype of Four Seasons (Figure 4). A 60-minute remote feedback session was conducted via Zoom with 5 patients with PsA from Portugal, the United Kingdom, the Netherlands, and Greece, using screen sharing for gameplay and an embedded SUS questionnaire via Google Forms. Similar to the cocreation session, participants were recruited through targeted outreach within the iPROLEPSIS network, through clinician referrals and patient advocacy networks affiliated with the project, using email and phone outreach. Participation was based on voluntary engagement and without the provision of compensation. For patients with PsA, exclusion criteria included cognitive impairments, severe visual/motor limitations preventing touchscreen interaction, or the inability to participate in digital sessions. The feedback session was also moderated by the last author, who did not participate in the quantitative analysis to minimize bias in the findings.

Data Collection

Cocreation Session

Audio recordings of the cocreation session were transcribed verbatim and then pseudoanonymized to preserve participant confidentiality. The digital setting facilitated active engagement from all attendees. The session was led by a female researcher specializing in qualitative research (last author), who guided discussions by posing questions and fostering group interactions to gather participants' opinions and experiences. To enhance participants' understanding of the session topics, relevant storyboards for each NoPain game were presented (Figure 3). The session followed a semistructured script with key questions designed to explore participants' views on the design of the proposed NoPain Games (Multimedia Appendix 1). A 90-minute timeframe was allowed for in-depth discussions and thorough exploration of the main topics.

Figure 3. Screenshot of the storyboards for the proposed NoPain games: (A) The “Lightning” game illustrates a memory challenge where players light up rooms to replicate a specific pattern; (B) The “Flashing Stars” game presents a relaxation activity focused on counting falling stars; and (C) “The Garden” game features a calming experience in which players identify specific flowers in a vibrant field.

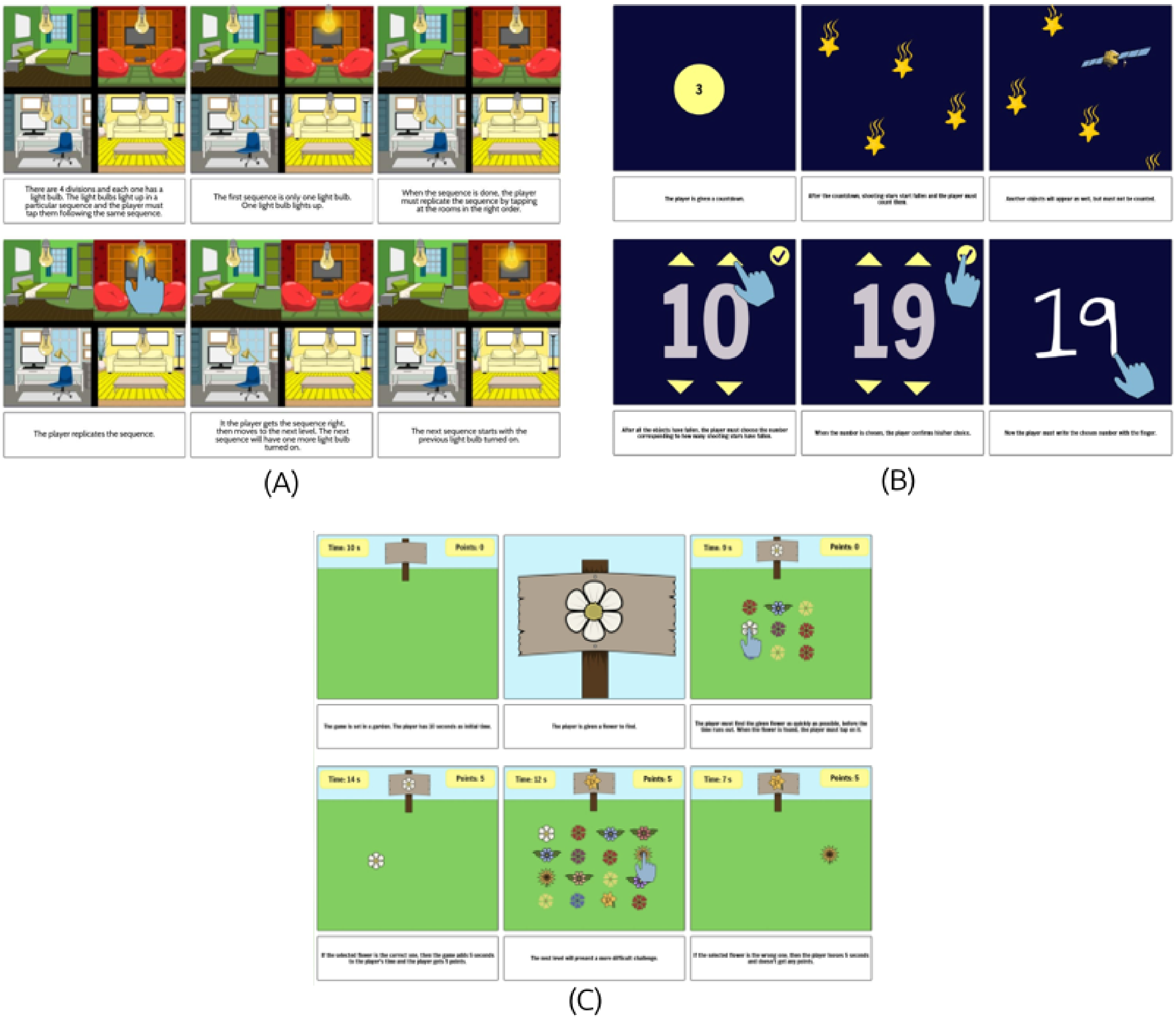
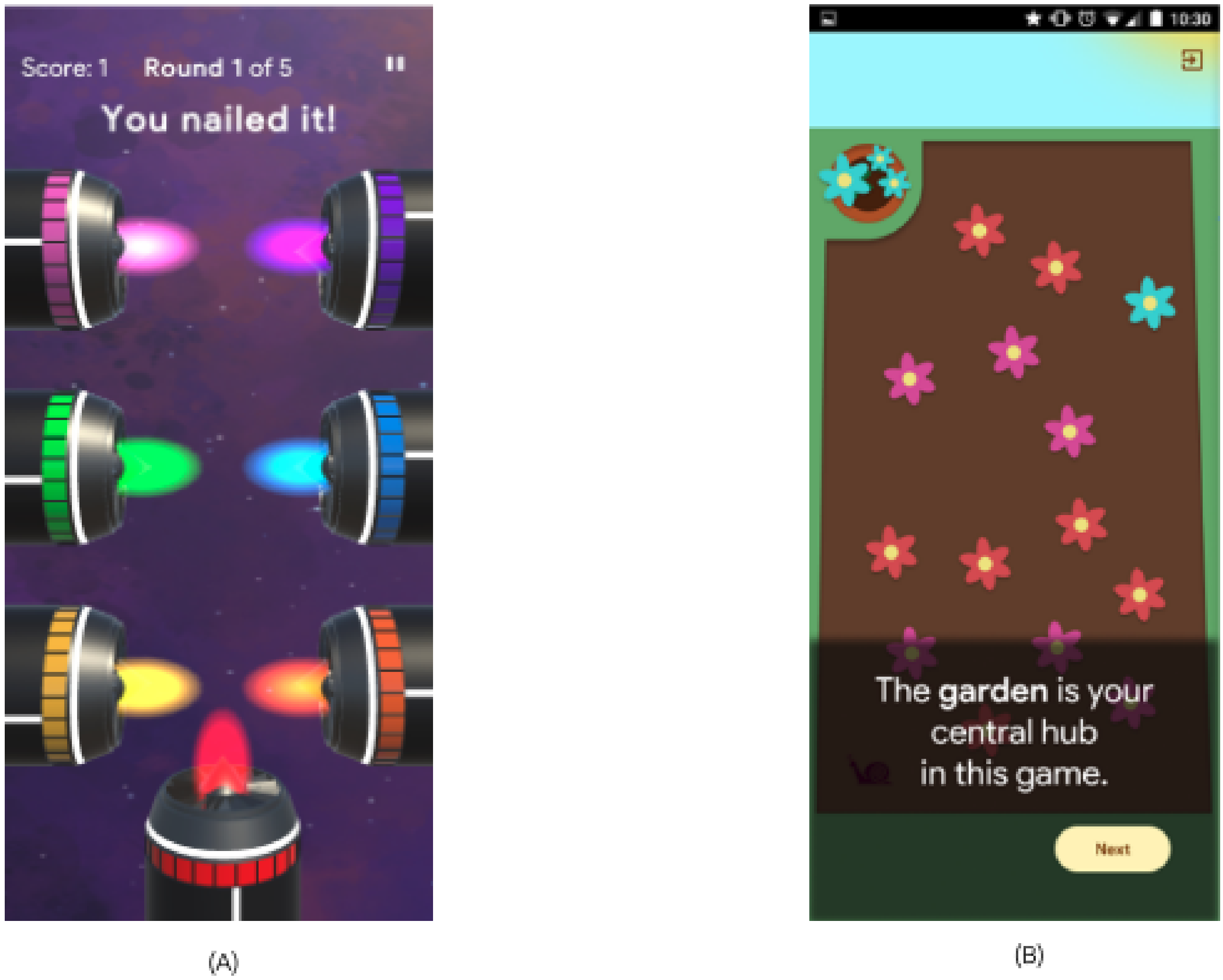


Figure 4. Screenshot of the prototypes used in the usability evaluation, featuring (A) the “Space Oddity” rockets minigame and (B) “The Four Seasons” design prototype.



Feedback Session

At the feedback session, after the playing experience and discussion, participants completed a System Usability Score (SUS) questionnaire [30,31], and verbally provided their feedback on refinements for improving game experience and accessibility. SUS comprises 10 statements, each evaluated on a 5-point Likert scale ranging from “Strongly Disagree” to “Strongly Agree.” It serves as a standardized tool for assessing user experience and pinpointing usability challenges. Widely adopted across academic and industry settings, the SUS has been validated as a reliable benchmark for interface evaluation [32]. Final scores range from 0 to 100, with values above 68 and above 80.3 generally indicating acceptable and excellent usability, respectively [32].

Table 1 tabulates the Implementation Matrix of the study. In the latter, qualitative and quantitative data sources are systematically organized by type, timing of data collection, participant group, and associated research aims. The matrix also outlines anticipated outcomes for each data stream, facilitating transparency in methodological integration and supporting the interpretive logic of the mixed methods design.

Table . Implementation matrix: overview of qualitative and quantitative data sources, collection timelines, participants, and study aim.

Data source	Type of data	Timing of collection	Participants	Study aim or research question	Anticipated outcome
Expert cocreation session	Qualitative	October 2023	14 experts (6 rheumatologists, 4 researchers, 4 technical experts)	What therapeutic priorities and design requirements should guide the development of NoPain Games?	Identification of core themes: therapeutic benefits, game difficulty, accessibility
Transcribed session dialogs	Qualitative	Postsession (Oct 2023)	Same as above	How do experts perceive the cognitive and physical needs of patients with PsA[a] in game design?	Thematic analysis yielding subthemes across cognitive and emotional engagement, gameplay adaptability, and inclusive interaction
Patient usability feedback session	Quantitative and qualitative	May 2025	5 patients with PsA (Portugal, the United Kingdom, Netherlands, Greece)	How usable and engaging are the NoPain Game prototypes from a patient perspective?	SUS[b] score, feedback on game mechanics, difficulty, and interaction design
SUS	Quantitative	May 2025	Same as above	What is the perceived usability of the NoPain Games?	Average SUS score >68
Patient feedback discussion	Qualitative	May 2025	Same as above	What refinements do patients suggest for improving game experience and accessibility?	Insights into difficulty adjustment, technical refinements, and inclusive interaction

[a]PsA: psoriatic arthritis.

[b]SUS: System Usability Scale.

Data Analysis

Qualitative Analysis

The qualitative data analysis process used an inductive thematic approach, with 2 independent researchers (first and last authors) analyzing all qualitative data. After thoroughly familiarizing themselves with the data, the researchers developed an initial set of codes, which were subsequently organized into themes and subthemes. These were carefully reviewed to ensure alignment with the data. The final themes and subthemes were crafted using clear and precise language, adhering to the methodology established by Braun and Clarke [33]. To enhance credibility and reliability, the themes underwent cross-validation, and agreement on code assignment was assessed using the Kappa statistic [34], with a threshold of 0.80 deemed acceptable [34]. The finalized themes were determined through consensus-building among the researchers. This approach provided a comprehensive understanding of the participants' perspectives, ensuring the analysis accurately captured their experiences and insights.

Quantitative Analysis

Quantitative data from the SUS questionnaire were analyzed with descriptive statistics, that is, mean, SD, median, IQR, showing how data spread out the central values, and 95% CIs quantifying the uncertainty in estimating the mean. Moreover, SUS 2D (average) and 3D (per patient with PsA) radar plots were estimated along with the SUS ICCs. The ICC is a foundational concept in item response theory, that is, a statistical framework used to model how individuals respond to test items based on an underlying trait or ability [32]. An ICC is a curve that shows the probability $p(\theta)$that a person with a given level of a latent trait θ (eg, usability perception, ability, attitude) will agree with or correctly answer a specific item. The probability $p(\theta)$is given by a 2-Parameter Logistic (2PL) model [35] as

(1)p(θ)=11+e−a(θ−b)

where a denotes the discrimination (slope), b the difficulty (threshold), and θ the latent trait value. A steep ICC (high a) means the item is highly sensitive to changes in usability perception, that is, ideal for detecting subtle shifts. A right-shifted ICC (high b) means the item is harder to agree with, that is, it activates only at higher usability levels. Finally, a flat ICC (low a) suggests the item does not discriminate well, that is, users across trait levels respond similarly. The whole quantitative analysis was carried out in Matlab 2025a (The MathWorks, Natick, USA).

Integration of Findings

Integration of qualitative and quantitative findings occurred through a sequential interpretive process that linked expert-derived design insights to patient-centered usability evaluation. Specifically, themes identified during the expert cocreation session—such as the need for cognitive stimulation, stress reduction, and adaptable difficulty—directly informed the development of 2 new game prototypes (Space Oddity and Four Seasons in the Results section). These prototypes operationalized therapeutic priorities and interaction mechanics discussed during the qualitative phase. In the subsequent feedback session, patients with PsA interacted with these games and provided both quantitative usability ratings via the SUS and qualitative feedback through moderated discussion. The SUS scores offered a standardized measure of perceived usability, while the qualitative comments contextualized these scores by highlighting specific design strengths (eg, intuitive controls, calming visuals) and areas for refinement (eg, onboarding clarity, difficulty pacing). Item-level analysis using ICCs further revealed how individual SUS items aligned with patient-reported experiences, enabling cross-validation of usability dimensions such as learnability and confidence (see Results). This integration strategy ensured that the design decisions were not only expert-informed but also empirically validated through patient engagement, reinforcing the iterative and user-centered nature of the development process.

Ethical Considerations

This study involved 2 distinct data collection sessions conducted under the iPROLEPSIS Horizon Europe project. They involved members of the iPROLEPSIS Consortium participating in the 2 sessions, complying with the iPROLEPSIS-received ethical approval from the Ethics Committee of Erasmus Medical Center, Rotterdam, Netherlands (MEC-2023‐0470). The first session was a cocreation workshop with members of the iPROLEPSIS Consortium, invited in their professional capacity. The second session was a usability feedback study involving patients with PsA. Neither session involved clinical procedures or the collection of sensitive personal health data. Participation in both sessions was entirely voluntary. For the cocreation session, verbal informed consent was obtained prior to participation. For the usability feedback session, informed consent was obtained via an embedded web-based form presented on the introductory page of the SUS questionnaire (Multimedia Appendix 2), which participants reviewed and accepted before proceeding. In both cases, participants were not compensated for their involvement. To ensure privacy and confidentiality across both phases, all audio recordings were transcribed verbatim and subsequently pseudo-anonymized. SUS survey responses were collected anonymously, with no identifying information linked to individual responses. All data were handled in accordance with General Data Protection Regulation–compliant data protection standards. No personally identifiable individuals appear in any images or supplementary materials.

Results

Qualitative Analysis

The cocreation session consisted of 14 participants, including 6 health care professionals, 4 technical experts, 3 observers, and one facilitator. The rheumatologists, all female, had a mean age of 36.3 (SD 7.2) years and over 10 years of professional experience, although most reported a basic level of technology literacy. The technical experts, all male, had a mean age of 44.8 (SD 7.9) years, with more than 15 years of experience and advanced-level technology literacy. The researchers, whose mean age was 28.5 (SD 7.0) years, included 3 females, each with over 4 years of experience and advanced-level technology literacy.

Thematic Analysis

Overview

Thematic analysis across all resolution scales, including themes and subthemes, revealed an almost perfect level of agreement between the 2 independent researchers (first and last authors), as indicated by a Kappa statistic of 0.90. In cases of uncertainty, discussions between researchers led to revisions, ultimately achieving 100% consensus across all subthemes. Data saturation was reached during the final stages of the cocreation session, as no new themes or subthemes emerged despite continued participant engagement. The structured yet open-ended nature of the sessions enabled iterative exploration of perspectives, and the consistency of responses across diverse stakeholders indicated that the thematic landscape had been sufficiently mapped. This saturation supports the robustness and completeness of the qualitative findings. From the thematic analysis, 3 primary themes emerged:

- Therapeutic benefits (Theme 1): This theme included subthemes such as pain distraction, memory enhancement, cognitive stimulation, stress reduction, and creative engagement, emphasizing the games’ potential to help patients with PsA manage both mental and physical challenges.
- Game difficulty (Theme 2): This theme centered on subthemes like balancing gameplay duration, complexity, and adaptive difficulty levels to sustain engagement while avoiding fatigue or diminished therapeutic impact.
- Accessibility and interaction (Theme 3): This theme highlighted subthemes addressing physical limitations, optimizing touchscreen usability, and implementing inclusive design to meet the specific needs of patients with PsA.

The key findings for each identified subtheme are summarized below.

Theme 1: Therapeutic Benefits

Overview

The therapeutic value of NoPain Games was a recurring topic during the discussions, underlining their role in addressing both physical and mental health challenges. Overall, the NoPain Games were admired not only for their ability to distract patients from pain but also for enhancing cognitive functions like memory. This dual benefit positions them as a promising digital intervention, especially for patients with PsA, who often face fatigue and mental health issues.

Subtheme 1: Cognitive Stimulation, Stress Reduction, and Pain Distraction

Specifically, participants emphasized the dual benefits of NoPain Games in providing cognitive stimulation and serving as effective pain distractors. Engaging in memory-focused activities not only addresses the psychological burden of chronic illnesses but also shifts focus away from physical discomfort, ultimately promoting holistic well-being. In particular, the unique appeal of the “Lightning” NoPain Game exemplifies these combined benefits, as one clinician noted:

> *I'm just thinking the 'Lightning' game is really interesting in the way that besides distracting, it's also a memory game […] because our patients are depressed, and this can be very helpful.* [Rheumatologist #2]

This sentiment highlights the potential of such games to uplift patients' moods and combat depression, showing the therapeutic value of cognitive engagement. Some participants further reinforced this aspect by stating,

> *I like this game [the 'Lightning' game], it promotes memory skills, which can be seen as pain distractors.* [Rheumatologist #3]

> *[…] the underlying mechanisms for this game [The Garden], is the memory, the mental stimulation in a way as part of the pain distractor.* [Researcher #2]

By promoting cognitive stimulation, these serious games can also contribute indirectly to pain management, reducing the focus on discomfort. Additionally, many participants acknowledged these games as powerful tools for pain distraction and stress reduction. One participant summarized this aspect succinctly, as follows:

> *The NoPain Games are intended to act as a pain distractor, allowing the patient to focus on it as a means of distraction, which in turn can support pain management and stress reduction.* [Researcher #1]

Overall, NoPain Games intends to integrate calming experiences and mentally stimulating activities, offering therapeutic benefits that extend beyond traditional treatments. By providing a distraction from chronic illness discomfort, these games shift patients' focus to engaging tasks that can alleviate stress, improve mood, and foster a sense of achievement. Their holistic approach addresses cognitive, emotional, and physical well-being that can enhance therapy outcomes and empower patients with PsA to actively participate in their care journey.

Subtheme 2: Promoting Creativity

This subtheme explores the integration of creativity into gameplay as a strategy for enhancing both physical and mental engagement in therapeutic contexts. Incorporating creative elements, such as drawing flowers in The Garden game, can introduce a unique combination of movement and creativity, directly supporting therapeutic objectives. As highlighted by one clinician:

> *Incorporating creative elements, such as drawing flowers, combines movement with creativity, stimulating the mind and physical tasks. Maybe the patient could draw the flower in this game [The Garden]. So you can combine in this way the movement and the creativity.* [Rheumatologist #6]

Furthermore, another clinician proposed expanding this creative aspect, as follows:

> *[…] maybe there is a stage where somehow someone can use this kind of flowers to build something […] more constructive.* [Rheumatologist #5]

Incorporating creative tasks into gameplay can potentially enhance interaction and personalization, encouraging active patient engagement in rehabilitation. Activities like drawing or constructing can stimulate fine motor skills, hand-eye coordination, cognitive focus, and emotional well-being. This creative emphasis can expand the NoPain Games' role, making them valuable tools that complement traditional treatments while combining therapeutic movement with self-expression for a more meaningful and enjoyable therapeutic experience.

Theme 2: Game Difficulty

Overview

Balancing the duration and complexity of gameplay emerged as a crucial aspect of game design during the cocreation session. Clinicians expressed concern that both extremes—prolonged sessions and overly simplistic mechanics—could undermine the games' therapeutic and engagement potential.

Subtheme 1: Managing Fatigue and Engagement

Prolonged gameplay or repetitive motions might exacerbate fatigue, reducing the games' effectiveness as therapeutic tools. One clinician pointed out the following:

> *For instance, it could be interesting to know how long the game sequences of the game [The 'Lightning' game] are, so how long should it last.* [Rheumatologist #5]

This indicates the need for carefully tailored session lengths to prevent overstimulation. On the other hand, overly simple game mechanics might fail to captivate players or provide meaningful stimulation. In this line, one clinician also highlighted this issue, noting:

> *So, I'm just thinking about the complexity of the game [The 'Lightning' game], in particular the duration of the sequence and the number of repetitions to play the game. […] have you decided or thoughts regarding the game duration?* [Rheumatologist #1]

This perspective highlights the need for future work to focus on designing gameplay that achieves the optimal balance—sufficiently challenging to maintain engagement, yet not so demanding as to provoke frustration or exhaustion.

Subtheme 2: Customizable Difficulty Levels

This subtheme introduces customizable difficulty levels and adjustable session durations as effective solutions. These features enable players to adapt gameplay to their individual needs, thereby enhancing both satisfaction and therapeutic outcomes. Furthermore, the significance of gradual progression in difficulty is emphasized by one participant:

> *For each level of difficulty, different coordination movements can be incorporated to offer a light yet progressively challenging experience, possibly linked to timing.* [Researcher #1]

This highlights how incorporating varied coordination movements fosters engagement while simultaneously supporting therapeutic goals, such as enhancing refined motor skills and cognitive reaction time. By offering challenges that escalate at an adaptable pace, players can feel a sense of growth and accomplishment, reinforcing motivation and adherence to therapeutic activities. Additionally, incorporating a time-based element was suggested as a potential customization option. As noted by one technical expert:

> *Can I also ask if the game [The 'Lightning' game] is connected to time? Does it involve reaction time that progressively get faster? For example, if there's a time limit, would it require completing the task within 5 seconds, then 10 seconds?* [Technical expert #1]

Together, these insights can suggest that customization in difficulty and session dynamics is vital for creating a flexible and impactful game experience.

Theme 3: Accessibility and Interaction

Overview

Accessibility was identified as another pivotal theme, reflecting the need for an accessible design to accommodate the physical limitations of patients with PsA. Symptoms like joint stiffness, swelling, and reduced dexterity can significantly impact interaction with digital tools, including games.

Subtheme 1: Physical Limitations and Interface Challenges

Overall, many clinicians highlighted specific challenges related to physical symptoms. For example, one clinician commented:

> *I guess if people have swollen fingers, the accuracy for this game ['The Garden' game] might be a problem, in the sense that you're trying to get quite a small area when you've got lots of flowers like this and the accuracy is potentially trickier.* [Rheumatologist #6]

This underlines the need for larger touch targets and simplified game interactions to accommodate patients with swollen fingers or a limited range of motion. Similarly, another clinician noted:

> *Sometimes, on a small screen, it's quite difficult to drag a marker to a particular point or to be very accurate with where we're pressing if we've got bad hand arthritis.* [Rheumatologist #2]

This observation points to the importance of designing user interfaces (UIs) that minimize the need for precise movements, making the games more accessible to those with severe joint issues.

Subtheme 2: Enhancing Inclusivity

To make the games more inclusive, developers could implement features such as adjustable touch sensitivity, alternative input methods (eg, voice commands), and options to magnify specific screen elements. These adaptations would enhance usability and ensure that the games can be enjoyed by a wider range of patients, regardless of their physical limitations.

These features foster a more inclusive experience for users with diverse needs, particularly those with conditions impairing their motor skills or visual acuity. For instance, adjustable touch sensitivity would allow users with reduced dexterity to interact with the interface more comfortably, while alternative input methods, such as voice commands, could provide a hands-free option for individuals unable to perform precise touch gestures. Similarly, magnification options for specific screen game elements could support users with visual impairments, ensuring they can navigate the game environment with ease and confidence. As suggested by one participant:

> *In fact, the levels of difficulty and synchronization with the user profile, as suggested before using the adaptation algorithm of the games (…) includes the patient's disease status, and eventually align with the hand or finger exhibiting better conditions in terms of symmetry and functionality.* [Researcher #1]

This highlights the potential for an adaptive system that tailors the gameplay experience to the specific physical conditions of the player. By considering factors such as hand symmetry, finger functionality, or disease progression, the game could dynamically adjust its mechanics to align with each patient's capabilities. For example, if a player has difficulty using one hand due to PsA, the game could prioritize interactions that favor the less-affected hand or allow for single-hand play. This ensures an engaging experience, minimizes frustration, and enhances therapeutic outcomes. Furthermore, a robust adaptation algorithm could learn and evolve with the player's needs, offering more personalized solutions over time by focusing on inclusive design principles.

Summarizing, these detailed insights into each theme emphasize the multifaceted nature of NoPain Games' design and impact. By addressing therapeutic benefits, gameplay balance, and accessibility, these games have the potential to make a meaningful difference in the lives of patients with PsA. Although the feedback from this session highlighted the clinical value of the storyboard elements and provided suggestions for design enhancements, there remains an opportunity to develop prototypes that integrate all these ideas to guide future development. The storyboards were revised based on the feedback collected, and a wider range of game mechanics was incorporated to improve the overall design and gameplay. Consequently, 2 NoPain game prototypes—Space Oddity Game and Four Seasons Game—have been conceptualized, as detailed next.

Prototyping the NoPain Games

Overview

Building on the subsequent phase of the agile approach after the ideation and storyboard design phase, 2 NoPain Games prototypes for pain relief were developed, incorporating targeted feedback gathered during the cocreation session (Development phase in Figure 1). These prototypes include Space Oddity (Figure 5), a space-themed serious game featuring 3 minigames that engage players in memory, coordination, and visual perception tasks; and Four Seasons (Figure 5D), a nature-themed game designed to promote repetitive, soothing movements. Tables 2 and 3 provide an overview of the key characteristics of the proposed NoPain Game prototypes, including their game concepts, clinical value, visual design, game mechanics, and difficulty progression.

Figure 5. Refined storyboards for the proposed NoPain Games: (A) Space Oddity coordination challenge, (B) Space Oddity memory challenge, (C) Space Oddity visual perception challenge, and (D) Four Seasons.

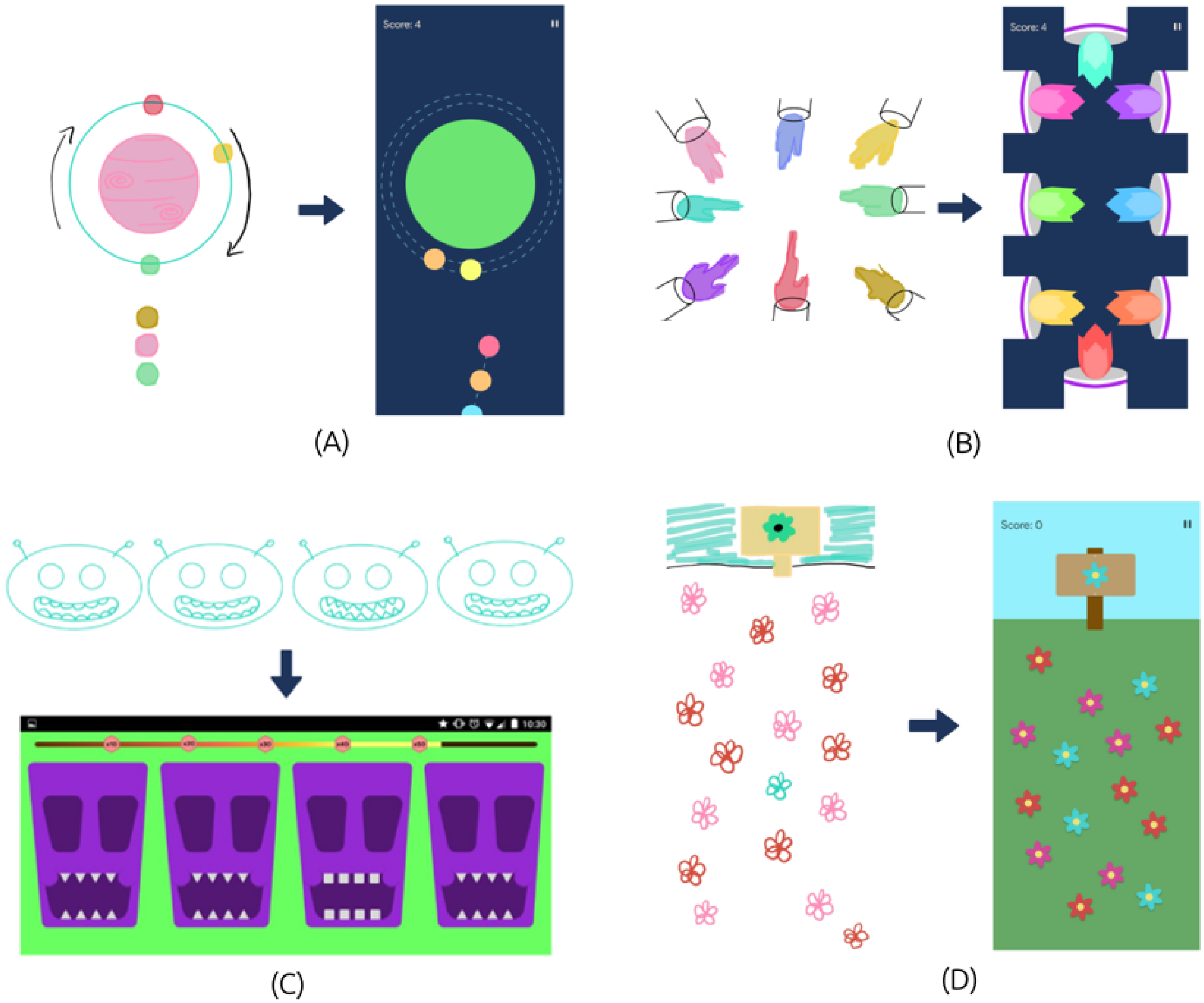


Table . Main characteristics of the proposed Space Oddity game prototypes, featuring 3 distinct mini-game–related areas, that is, memory-based challenge, coordination skills, and visual perception, each designed to target specific cognitive and motor functions associated with psoriatic arthritis. For each area, the game concept, intended clinical value, visual design elements, core gameplay mechanics, and difficulty progression are described.

	Memory-based challenge	Coordination skills	Visual perception
Game concept	Players memorize and reproduce a sequence of ignited spaceships.	Players launch asteroids into orbit while avoiding collisions.	Players identify the odd creature among a group.
Clinical value	Strengthening cognitive and memory sequencing skills.	Developing motor coordination by improving reaction timing and precision.	Improving focus, visual perception, and cognitive processing speed.
Visual design	Embracing the game's space esthetic, the spaceships' ignition includes vibrant and distinct hue colors to strengthen visual memory cues during gameplay, with clear contrast against the background for easy recognition.	A central, colorful planet with orbiting asteroids stands out against a calm-toned universe, with new asteroids emerging from below to ensure an intuitive and structured gameplay flow.	Aligned with the futuristic theme, alien robot-like creatures were chosen as the main characters, since various features (such as eye color, teeth shape, and accessories) can be easily modified for a dynamic experience.
Game mechanics	At the start of each round, a sequence of rockets lights up. The player must replicate the sequence by tapping the rockets in the correct order. The game verifies the input, granting progression if correct. If incorrect, the sequence repeats for the player to try again.	The game consists of multiple stages, each with a set number of asteroids. Players must launch asteroids into orbit while ensuring they do not collide. When all asteroids are placed successfully, a new planet appears. If 2 asteroids collide, the game ends.	A set of creatures is displayed, with one differing in subtle ways. The player must identify the intruder. Feedback is given through color and audio cues.
Difficulty progression	The sequence length increases as the game progresses, making it more challenging. An easier mode maintains the sequence and adds only one new element per round, while a harder mode randomizes the sequence each time, increasing difficulty.	Initially, asteroids move in a straight path toward the orbit. At higher difficulty levels, additional asteroids move dynamically, requiring more precise timing. Some areas on the planet may be restricted (eg, mountains), and color-coded zones may require matching asteroids.	At lower difficulties, the creatures' differences are more apparent. At higher levels, the distinctions become more subtle, requiring greater attention to detail.

Table . Main characteristics of the proposed Four Seasons Game prototypes, targeting 2 distinct areas, that is, visual perception and coordination skills, each designed to support stress reduction, focus, and motor precision in individuals with psoriatic arthritis. For each area, the game concept, intended clinical value, visual design elements, core gameplay mechanics, and difficulty progression are described.

	Visual perception	Coordination skills
Game concept	Players identify specific flower species on a garden field.	The player has to select fruits to fall from a tree, while ensuring they fall into a basket and not on the floor.
Clinical value	Stress reduction, improved focus, and coordination.	Reaction time and precision.
Visual design	A relaxing nature-based theme in the form of a garden. Various colorful plants and flowers populate the garden, giving it a vibrant atmosphere.	A nature-based theme with fruit trees in the background full of colorful fruits.
Game mechanics	The player is instructed to find and tap a chosen flower species, in a garden scenario full of various plants. Feedback is provided to let the player know if they selected the correct flower. As the game progresses, the seasons change, creating a dynamic environment and increasing the level of challenge.	The player has to select fruits to fall from a tree, as a rolling bucket passes underneath. The player gains points if they correctly calculate the timing of the falling fruit, making them land in the basket. They lose points if the fruits miss the basket instead.
Difficulty progression	As the game progresses, some elements become more challenging. For example, the flowers are no longer static; they start moving or become more similar to each other.	As difficulty rises, the number of fruits increases, the basket moves faster, and the basket becomes smaller.

Space Oddity

The Space Oddity NoPain Game was inspired by the space theme expressed in Figure 3B and incorporated a memory-based challenge with a patterns-matching mechanic similar to that in Figure 3A. Based on clinician feedback provided at the second design phase, other game tasks related to coordination and visual perception were added, while maintaining the overall space thematic and base mechanics. Table 2 links the specific game tasks with the themes highlighted at the second cocreation session. Throughout the game, players should experience all these tasks in the form of minigames. Therefore, as the players enter Space Oddity and select a difficulty level, 3 minigames appear in a different order to avoid monotony and ensure that the game will be played various times without getting repetitive.

The first minigame is designed to train coordination skills (Figure 5A). A main planet stands at the center of the screen, while being surrounded by a fixed-speed orbit of asteroids. As new asteroids emerge from below, the player/patient must launch them into orbit one at a time, ensuring they do not collide with existing asteroids. Success depends on precise timing, requiring the player/patient to tap the screen at the right moments to avoid collisions. After a certain number of asteroids are launched, a new planet emerges with new challenges. Additional features to add diversity to the game include areas on the planet with mountains where asteroids cannot be placed, or color-coded zones to correspond to asteroids of the same color (Figure 5A). The second minigame, depicted in Figure 5B, features 8 spaceships scattered across the screen, all starting with their engines turned off. It operates, like the Simon game [36], in rounds where players watch as a sequence of rockets ignite and then are required to replicate the sequence by selecting the rockets in the correct order. The first round begins with a single randomly ignited rocket. In each subsequent round, the previous sequence repeats in the same order, with one additional rocket added. The minigame continues until a certain number of rounds is successfully completed. To enhance memorization, unique colors and musical notes can be attributed to each rocket, providing both visual and auditory stimuli. Figure 5B illustrates the core prototype elements and an example screen in which all rockets are ignited with their distinct colors. The third minigame focuses on memory, concentration, and visual perception, where a player/patient performs several rounds of identifying from a set of 4 creatures, the one that differs from the others (intruder) (Figure 5C). Overall, various features may be modified to create distinctions, such as the creatures' shape or variations in their elements.

Four Seasons

The Four Seasons NoPain Game builds upon the concepts illustrated in the storyboard of Figure 3C. Similarly, Tables 2 and 3 link the specific game tasks with the themes highlighted at the second cocreation session. During gameplay, the players/patients explore a garden filled with various flowers and are asked to identify a specific one (Figure 5D). As the game progresses, the seasons change, affecting both the environment's aesthetic and the challenges presented. Each season introduces unique difficulties and advantages, adding variety to the game. The Four Seasons game was designed to distract players from pain by offering 2 distinct modes to suit individual preferences. The Relax Mode allows players to peacefully search for flowers in a garden across different seasons, while the Party Mode introduces added excitement with minigames at the end of each season for a more dynamic experience. Overall, this game, designed to reflect the unique characteristics of each season, has the potential to offer a refreshing gameplay variety while introducing an element of dynamism for players/patients seeking a more immersive experience.

Feedback From Patients With PsA

Quantitative Analysis

The patients with PsA feedback group included 2 women and 3 men, with ages ranging from 25 to 64 years, split into age groups of 25 - 34 years (1), 35 - 44 years (1), 45 - 54 years (2), and 55 - 64 years (1). All patients had been diagnosed with PsA for more than 8 years; had an average experience in gaming, and good experience in smartphone usage. During the session, 4 participants (1-4) reported experiencing a PsA flare.

The session first focused on "Space Oddity" through a playable version developed in Unity3D (Figure 4A). Participants played the game and then evaluated its usability by completing the SUS questionnaire. Table 4 tabulates the derived descriptives from the SUS data, that is, mean (SD), 95% CI, and median (IQR). From the latter, it is seen that an average SUS score of 79 (SD 10.4; 95% CI 69.89 - 88.11) was reached, suggesting good overall usability, with some variation in individual user perceptions.

Table . Descriptives of the System Usability Scale (SUS).

Item definitions	Mean (SD)	95% CI	Median (IQR)
Frequency Use (1)	3.80 (0.84)	3.07 - 4.53	4 (3-4)
Complexity (2)	2.40 (1.95)	0.69 - 4.11	1 (1-4)
Ease of Use (3)	4.60 (0.55)	4.12 - 5.08	5 (4-5)
Support Needs (4)	1.00 (0.0)	1.00 - 1.00	1 (1-1)
Functionality Integration (5)	3.80 (0.84)	3.07 - 4.53	4 (3-4)
Inconsistency (6)	2.00 (1.0)	1.12 - 2.88	2 (1-3)
Confidence (7)	4.40 (0.55)	3.92 - 4.88	4 (4-5)
Cumbersomeness (8)	1.60 (0.90)	0.82 - 2.38	1 (1-2)
Learnability (9)	4.20 (0.84)	3.47 - 4.93	4 (4-5)
Learning Curve (10)	2.20 (1.64)	0.76 - 3.64	2 (1-2)
Total SUS	79 (10.4)	69.89 - 88.11	75 (75 - 82.5)

Moreover, Figure 6 displays a 2D radar plot of average SUS responses across all participants, with each axis representing a usability factor mapped to its corresponding SUS item number (eg, Frequency Use (1), Cumbersomeness (8), etc). The radial layout allows for intuitive comparison of agreement levels across items, with values ranging from 1 (“Strongly Disagree”) to 5 (“Strongly Agree”). Peaks in the plot are observed around positively worded items such as Ease of Use (3), Confidence (7), and Learnability (9), indicating strong agreement and direct contributions to higher SUS scores. Conversely, troughs appear around negatively worded items like Support Needs (4), Cumbersomeness (8), and Learning Curve (10), where low agreement is desirable and reflects positive usability perceptions. The overall shape of the radar plot reveals a balanced usability profile, with high agreement on core interaction elements and consistent disagreement with statements implying complexity or poor integration.

In addition, Figure 6B presents a 3D radar surface plot where axes radiate outward from the center, each labeled with a usability factor and its corresponding SUS item number, for example, Consistency (6) and Learning Curve (10). The surface height represents participant index, while the color gradient (blue to yellow) encodes response values from 1 to 5. The plot reveals structural patterns in usability perception, with elevated regions around Ease of Use (3), Confidence (7), and Learnability (9), positively worded items where high agreement directly contributes to higher SUS scores. Conversely, lower scores on negatively worded items such as Complexity (2), Support Needs (4), Inconsistency (6), Cumbersomeness (8), and Learning Curve (10) are desirable, as they reflect disagreement with statements implying poor usability. Notably, Support Needs (4) received a uniform score of 1 across all participants, indicating unanimous disagreement with the notion that external support was required to use the system, an encouraging signal of intuitive design and self-sufficiency. Participant-level differences are also evident. In particular, Participants 1 - 4, who reported experiencing PsA flare episodes during evaluation, showed slightly more variability in items such as Functionality Integration *(5)* and Inconsistency *(6)*, suggesting that symptom severity may influence perceptions of system responsiveness and coherence. In contrast, Participant 5, who did not report a flare, exhibited consistently high agreement on positively worded items and low agreement on negatively worded ones, resulting in a smoother and more elevated usability profile across the radar surface. Together, Figure 6A and B offer complementary perspectives on usability perception, highlighting both structural consistency and individual-level nuances.

Figure 6. (A) Radar chart of mean SUS item scores arranged clockwise, highlighting usability strengths and weaknesses. (B) 3D radar surface plot of individual SUS responses, showing participant-level variation across usability dimensions. SUS: System Usability Scale.

Furthermore, Figure 7 illustrates a detailed visualization of item-level behavior across the latent usability trait spectrum. The left panel presents 2D ICCs for selected SUS items, each labeled by item index. These curves illustrate the probability of agreement p(θ) (see (1)) as a function of latent usability (θ), revealing distinct patterns of discrimination and difficulty. The right panel complements this view with a 3D ICC ribbon plot, where each SUS item is vertically separated and labeled by its full descriptor. This format enhances interpretability by exposing how response probabilities vary not only across θ but also across items. Moreover, Table 5 includes the corresponding parameters (a,b) (see (1)) of the ICCs depicted in Figure 7. Analytically, from Figure 7 and Table 5, the following observations can be derived per SUS item, revealing their heterogeneity:

- Frequency of Use (1)-positive wording: With a = 1.29 and b = –1.10, this item shows moderate discrimination and early activation. The curve is gently sloped and elevated across the trait spectrum, indicating consistent agreement even at low usability levels. This suggests it reflects habitual or contextual familiarity rather than sensitivity to usability changes. Its contribution is stable but diagnostically limited.
- Complexity (2)-negative wording: With a = 0.67 and b = 0.49, the curve rises gradually and activates in the mid-range of θ. Users increasingly disagree as usability improves, aligning with expected polarity. However, the low a value indicates weak discrimination. The ribbon is broad and shallow, offering general friction-related information but lacking precision.
- Ease of Use (3)-positive wording: Despite its conceptual importance, a = 1.73×10^{13} and b = $–5.93\times10^{14}$ suggest a modeling anomaly, likely due to uniform agreement across participants. The curve appears flat and saturated, with no meaningful slope. While users universally endorse this item, its statistical discrimination is negligible, making it confirmatory rather than diagnostic.
- Support Needs (4)-negative wording: With a = $–8.72\times10^{14}$ and b = $–1.18\times10^{1}$, this item exhibits modeling instability. The curve is flat and extremely right-shifted, indicating that nearly all users disagreed regardless of usability level. In fact, all users disagreed. This uniform rejection aligns with high usability but limits the item’s ability to differentiate experiences. It may be conceptually relevant but statistically redundant.
- Functionality Integration (5)-positive wording: With a = –0.11 and b = 3.48, the curve is shallow. The negative a suggests an inverse slope, possibly due to inconsistent response patterns. The ribbon is low and delayed, indicating poor discrimination and limited responsiveness.
- Inconsistency (6)-negative wording: Like item 4, a = $–8.72\times10^{14}$ and b = $–1.18\times10^{1}$ reflect modeling collapse. The curve is flat and nondiscriminative, with uniform disagreement across users. While this aligns with high usability, the lack of slope or variability renders it statistically inert.
- Confidence (7)-positive wording: With a = 1.73×10^{13} and b = $–5.93\times10^{14}$, this item shows extreme early activation and saturation. The curve is flat and elevated, indicating universal agreement. Like *Ease of Use (3)*, it functions as a strong confirmatory item but lacks statistical discrimination due to the ceiling effect.
- Cumbersomeness (8)-negative wording: Again, a = $–8.72\times10^{14}$ and b = $–1.18\times10^{1}$ suggest modeling failure. The curve is flat and low, with near-universal disagreement. While this reflects low friction, the lack of variability limits its diagnostic value.
- Learnability (9)-positive wording: With a = 59.00 and b = –2.28, this item shows extremely high discrimination and early activation. The curve is steep and sharply contoured, making it the most responsive item in the set. The ribbon is narrow and elevated, indicating strong sensitivity to usability perception across the trait range. This item is statistically and conceptually robust.
- Learning Curve (10)-negative wording: With a = 0.33 and b = 4.33, this item exhibits low discrimination but great difficulty. The curve rises gradually and activates only at the upper end of the usability trait spectrum. The ribbon in Figure 7 is delayed but upward-sloping, confirming that the item captures late-stage usability clarity, particularly in systems that excel in onboarding. While its responsiveness is limited at lower θ levels, it still contributes meaningful coverage of the high-usability tail, making it useful for identifying systems with strong learnability.

From the aforementioned, we can identify four interpretive categories: (1) Statistically robust item: “Learnability (9)” shows exceptional discrimination and early activation, making it the most diagnostically powerful item. (2) Confirmatory but saturated items: ”Ease of Use (3)” and “Confidence (7)” are universally endorsed but exhibit near-zero discrimination due to ceiling effects. They confirm high usability but offer limited differentiation. (3) Low-performing or inert items: “Support Needs (4),” “Inconsistency (6),” and “Cumbersomeness (8)” show flat, nonresponsive curves and modeling collapse, limiting their diagnostic utility. (4) Mid-range and coverage items: “Frequency of Use (1),” “Complexity (2),” and “Learning Curve (10)” contribute moderate or delayed responsiveness, offering coverage across the trait spectrum but with limited precision. This item-level interpretation supports evaluation of the SUS, guiding decisions on item weighting and further scrutiny for potential refinement.

Figure 7. 2D Item characteristic curves (left) showing agreement probability p(θ) across latent usability trait (θ), with each curve labeled by item index, along with the 3D item characteristic curve ribbons (right) for all System Usability Scale items, visualizing response probabilities across θ and item index, with full item labels on the horizontal axis.

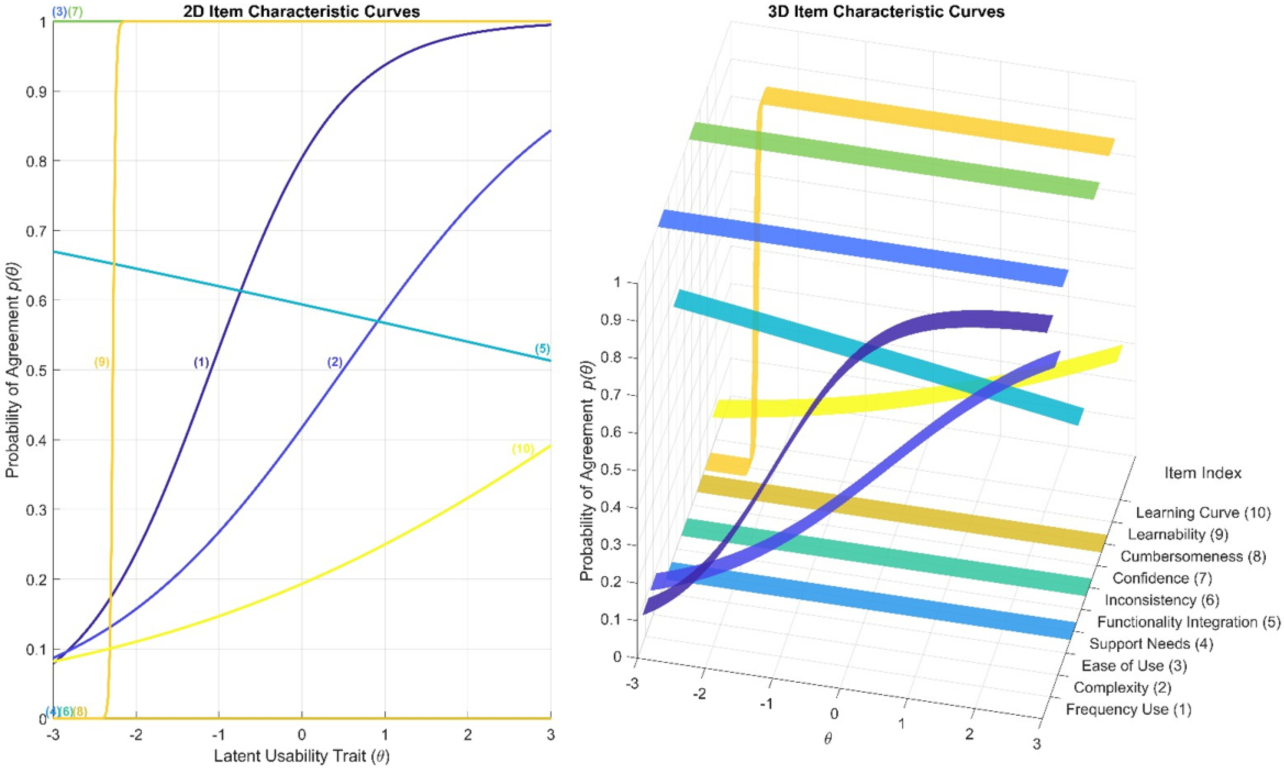


Table . Estimated parameters (a, b) of the corresponding item characteristic curves of Figure 7 for each System Usability Scale (SUS) item.

SUS item	Alpha (a)	Beta (b)
Frequency Use (1)	1.29	–1.10
Complexity (2)	0.67	0.49
Ease of Use (3)	1.73×10^{-13}	$–5.93\times10^{14}$
Support Needs (4)	$–8.72\times10^{-14}$	$–1.18\times10^{15}$
Functionality Integration (5)	–0.11	3.48
Inconsistency (6)	$–8.72\times10^{-14}$	$–1.18\times10^{15}$
Confidence (7)	1.73×10^{-13}	$–5.93\times10^{14}$
Cumbersomeness (8)	$–8.72\times10^{-14}$	$–1.18\times10^{15}$
Learnability (9)	59.00	–2.28
Learning Curve (10)	0.33	4.33

Qualitative Analysis

During the patients with PsA feedback session, the group was guided through a design prototype demonstration of the Four Seasons (Figure 4B), covering its tutorial, practice session, navigation, gameplay, bonus level, and scoring. Feedback was generally positive and constructive, with discussions focusing on several key areas across the 2 games. In particular, they proposed:

- resolving iOS compatibility issues to improve accessibility,
- modifying difficulty levels and incorporating adaptive artificial intelligence (AI) in "Space Oddity" to tailor challenges to individual player profiles,
- transitioning "The Four Seasons" from prototype to full interactive development with seasonal mechanics and high-quality 3D rendering, and
- introducing smartwatch integration to track stress levels before and after gameplay in both games.

From a combined perspective, the derived average SUS score of 79 reflects a generally positive user experience, placing the NoPain Games within the "good" usability range according to established benchmarks. This score suggests that core interaction elements, such as navigation, responsiveness, and interface clarity, were well-received across a diverse group of patients with PsA, even during flare episodes. To better understand how this overall score reflects item-level behavior, we examined the individual ICC profiles. The accompanying ICC analysis revealed several positive outcomes. These item-level patterns help contextualize the overall SUS score of the NoPain Games, revealing how specific usability dimensions contributed to the positive user experience. This is combined with the feedback provided during the session, which revealed that usability alone does not fully capture user expectations in therapeutic contexts. Participants articulated a clear desire for functional enhancements and personalization features that extend beyond baseline usability. Their suggestions, ranging from adaptive AI and platform compatibility to immersive rendering and biometric integration, indicate that patients are not merely evaluating ease of use, but actively envisioning how the system could evolve into a more responsive and clinically meaningful tool. This feedback reveals the importance of interpreting usability scores in tandem with qualitative insights, especially when designing interventions for chronic conditions where therapeutic relevance and emotional engagement are critical.

Discussion

Main Findings

This study aimed to explore expert perspectives on the design of 2 serious games, that is, NoPain Games, intended to alleviate chronic pain and improve quality of life for individuals with PsA. The cocreation session revealed 3 key thematic areas: therapeutic benefits, game difficulty, and accessibility and interaction. These themes directly informed modifications to the game prototypes, aligning with the study's objective to develop personalized, user-centered digital interventions for chronic pain management. Moreover, the usability assessments of patients with PsA using SUS and feedback further supported the design approach, indicating generally positive perceptions of the game prototypes and reinforcing their potential as accessible and engaging tools for digital pain management. The extended quantitative analysis of SUS responses, including item-level discrimination and difficulty modeling via ICCs, provided deeper insight into usability perceptions. These findings offer item-level insights that can guide targeted enhancements to interface design and interaction mechanics. This quantitative granularity complements the thematic findings, particularly those related to therapeutic intent and user engagement.

Therapeutic benefits were a major focus, with the games designed to provide pain distraction and cognitive engagement. For example, the "Lightning" game focuses on memory enhancement, which can help distract from pain and improve cognitive function [3]. Additionally, the games incorporate soothing and repetitive mechanics, such as in the "Flashing Stars" game, which promotes relaxation and stress reduction [6]. Moreover, previous studies had also suggested a link between serious games and improvements in working memory [37], prompting further research on this topic.

Game difficulty was another critical theme, with feedback emphasizing the need to balance game duration and complexity to maintain engagement without causing fatigue. This led to the development of adaptive difficulty levels that adjust to the player's abilities [38]. The games were designed to be challenging enough to keep players engaged but not so difficult that they become frustrating. This balance is crucial for maintaining the therapeutic benefits of the games [39]. In fact, the challenge of balancing game difficulty has been identified in previous studies [38,39], yet it remains complex due to the need to provide therapeutic benefits while maintaining player engagement. When designing serious games, it is essential to balance activity levels, as over-exercise can cause fatigue and frustration, while insufficient activity may reduce therapeutic benefits. Adaptive difficulty levels aligning with players' abilities and goals could ensure engagement, gradual improvement, and overall well-being.

Accessibility and interaction were also key considerations. Overall, the game design considered the physical limitations of patients with PsA, such as joint stiffness and swelling. This led to the optimization of touchscreen usability and the inclusion of larger touch targets to accommodate swollen fingers [40]. Ensuring that the games are accessible to all patients with PsA, regardless of their physical abilities, was a key focus. This included designing intuitive and easy-to-use interfaces [41]. Moreover, smartphone accessibility has been analyzed in various domains [40,41] due to the fine motor skills required for small touchscreen interactions. However, its impact on patients with PsA could be investigated further.

The alterations in game design during the cocreation workshops align with findings from existing literature on serious games and digital health interventions. Studies have shown that serious games can reduce stress, enhance cognitive function, and provide pain relief through immersive and engaging experiences [3,6]. The NoPain Games leverage these benefits by incorporating memory and relaxation activities. The importance of balancing game difficulty to sustain engagement and therapeutic efficacy is well-documented. Adaptive difficulty levels, planned for implementation in the NoPain Games, are crucial for maintaining player interest and ensuring the games' effectiveness [38,39]. Research highlights the need for accessible game design, especially for individuals with physical limitations. The NoPain Games' focus on optimizing touchscreen usability and inclusive design reflects these findings [40,41]. Furthermore, the conceptual framework (2D-ME) for explaining self-first and self-third person views of prototyping dynamics in serious games design highlights the importance of iterative feedback and dynamic constructs in the game design process [42]. This framework supports the iterative refinement seen in the NoPain Games development, ensuring that the games are both engaging and therapeutic.

Furthermore, the design of NoPain Games leverages key insights into how the human brain pays attention to various aspects of game design. By understanding and applying principles related to perception, memory, attention, and emotional engagement, the developers can create games that are both therapeutic and engaging.

For example, perception plays a crucial role in how players interact with NoPain Games. The brain processes visual and auditory information to understand and navigate the game world. For instance, the "Lightning" NoPain game uses effective visual cues and sound effects to guide players' attention and enhance their immersion [43]. The UI design is also critical, ensuring that players can easily access information and controls without being distracted by confusing or cluttered interfaces. This is particularly important for patients with PsA, who may have physical limitations affecting their game interaction [40].

Memory is another essential aspect of the proposed NoPain Games. Players rely on short-term memory to remember recent actions, objectives, and game mechanics. This is particularly important in fast-paced games where quick decision-making is required. For example, the “Lightning” NoPain Game focuses on memory enhancement, helping players to distract from pain and improve cognitive function [3]. Long-term memory is engaged when players learn and remember game rules, storylines, and strategies. By reinforcing learning through repetition and rewards, NoPain Games can enhance player retention and engagement [43].

Attention is also vital for maintaining focus and engagement in the proposed NoPain Games. The brain’s ability to maintain focus is critical for sustained engagement, and the developers have used various techniques to capture and hold players’ attention. This includes compelling narratives and dynamic gameplay [44]. Managing cognitive load is essential to prevent player fatigue. NoPain Games balance complexity and simplicity to keep players engaged without overwhelming them. This involves designing intuitive controls, clear objectives, and manageable challenges, ensuring that the games are accessible and enjoyable for patients with PsA.

Emotional and motivational factors significantly influence the brain’s attention to NoPain Games. Emotions play a significant role in how players experience the games. Engaging storylines, relatable characters, and emotional rewards enhance player immersion and satisfaction. The brain responds to rewards and achievements, which can motivate players to continue playing. NoPain Games use various reward systems, such as points and progression systems, to keep players motivated and engaged [43].

Achieving a flow state is another critical aspect of NoPain Games. The concept of “flow” refers to a state of deep focus and immersion where players lose track of time [45]. Achieving flow involves balancing challenge and skill, providing clear goals, and offering immediate feedback. NoPain Games successfully induce flow by creating a balance between challenge and skill, ensuring that players remain deeply engaged and enjoy the therapeutic benefits of the games [43].

Approaching the aforementioned from the lens of the 3 themes identified, that is, therapeutic benefits, game difficulty, and accessibility and interaction, a deep interwovenness can be identified with sustained engagement, player attention, cognitive load, player fatigue, emotions, and flow state in the context of NoPain Games for patients with PsA. In particular:

1. Therapeutic benefits: NoPain Games aim to alleviate pain and improve emotional well-being by providing cognitive stimulation, stress reduction, and memory enhancement. These benefits are crucial for fostering sustained engagement as players find therapeutic value in their activity, which keeps their attention fixed on the games. The soothing and predictable mechanics of games like “Flashing Stars” reduce *cognitive load* by presenting clear objectives and repetitive tasks, allowing players to focus without being overwhelmed. This emotional engagement also helps maintain the ideal flow state, where players are immersed in gameplay and distracted from their chronic pain.
2. Game difficulty: Balancing and adapting difficulty levels directly supports sustained engagement by keeping tasks manageable yet stimulating. Customizable difficulty prevents player fatigue, ensuring sessions remain enjoyable rather than exhausting. A dynamically adjusted challenge ensures that attention is maintained without inducing frustration, helping to optimize both players’ attention and cognitive effort. Gradual progression of difficulty, such as increasing complexity in the “Lightning” game, enhances emotions of achievement and satisfaction, which is key for reinforcing engagement and maintaining the flow state.
3. Accessibility and interaction: The design considerations in this theme, such as larger touch targets and intuitive interfaces, ensure that physical limitations do not hinder participation. This inclusivity supports sustained engagement by enabling patients to interact easily and prevent frustration. These accommodations reduce the cognitive load on players, allowing them to focus on the tasks rather than overcoming interface challenges. This theme also reduces players’ fatigue by minimizing physical strain and frustration, helping to maintain a relaxed state essential for emotional well-being and immersion into a flow state.

Together, these themes create a synergistic experience where accessibility removes barriers, therapeutic benefits uplift emotional and cognitive states, and adaptive difficulty sustains attention and motivation, all of which harmonize to engage patients with PsA while mitigating pain effectively.

The iterative optimization process in developing NoPain Games prototypes involved refining visual cues, simplifying the UI, balancing cognitive load, and incorporating emotional and motivational elements. By understanding how the brain pays attention to different aspects of game design, the designers/developers were able to create games that are both effective and enjoyable for patients with PsA. This approach ensured the potential for the proposed NoPain Games to provide therapeutic benefits while maintaining high engagement and satisfaction levels.

Overall, the cocreation themes guided the iterative refinement of the NoPain Games, confirming they are both engaging and therapeutic for patients with PsA. The integration of feedback from the cocreation session with insights from the relevant literature review resulted in 2 game prototypes that are well-suited to the needs of their target audience, as expressed via the feedback from patients with PsA. In fact, the patient feedback session extended the thematic framework by introducing concrete implementation priorities that reflect lived experience and technical expectations. The mean effective SUS score (79) confirmed good usability, yet the qualitative suggestions revealed areas for refinement that map directly onto the 3 core themes.

First, the recommendation to resolve iOS compatibility issues reinforces theme 3 (accessibility and interaction). While the expert session emphasized touchscreen optimization and inclusive design, patients highlighted platform-specific barriers that could limit access. This aligns with Beltran-Alacreu et al [16], who emphasized the importance of device-level accessibility in digital interventions for chronic pain populations. Addressing cross-platform compatibility is essential to ensure equitable access, particularly for older adults or those with limited digital literacy.

Second, the proposal to incorporate adaptive AI into Space Oddity to tailor difficulty to individual profiles directly advances theme 2 (game difficulty). This feedback moves beyond static balancing and introduces personalization as a therapeutic strategy. It resonates with Tuah et al [20], who identified progression systems and adaptive mechanics as key to sustaining engagement in rehabilitation games. In the context of PsA, where symptom severity and cognitive capacity fluctuate, AI-driven difficulty modulation could enhance both usability and therapeutic relevance.

Third, the suggestion to transition the Four Seasons from prototype to full interactive development with seasonal mechanics and high-quality 3D rendering reflects an evolution of theme 1 (therapeutic benefits). While experts emphasized cognitive stimulation and stress reduction, patients implicitly called for deeper immersion and aesthetic refinement. This aligns with Gromala et al [6], who argued that realistic environments and sensory richness contribute to pain relief. The seasonal metaphor may also support emotional regulation by anchoring gameplay in familiar, cyclical rhythms.

Finally, the idea of integrating smartwatch-based stress tracking introduces a novel extension to Theme 1, bridging subjective experience with physiological feedback. This aligns with recent findings by Pinge et al [46], who systematically reviewed wearable-based stress detection and highlighted the clinical potential of physiological signals such as heart rate variability and electrodermal activity for real-time stress monitoring. By capturing pre- and post-game stress levels, future iterations of NoPain Games could offer biofeedback-informed personalization, enhancing therapeutic precision and enabling longitudinal tracking within digital care ecosystems.

Overall, these patient-driven suggestions not only validate the thematic structure but also push its boundaries toward real-world deployment. They emphasize the importance of technical adaptability, personalized challenge design, and physiological integration in therapeutic game development. More broadly, they demonstrate how participatory feedback can transform conceptual frameworks into actionable design trajectories, ensuring that digital health tools remain grounded in both clinical insight and patient reality.

Limitations

Although this study offers important insights, it is essential to acknowledge certain limitations. The study engaged a limited sample of 14 experts from 3 European countries in one session, followed by 5 patients with PsA from 4 European countries in a subsequent session, including SUS-based evaluation. While the results offer initial insights into usability perceptions, the limited sample size may have constrained variability, leading to some of the ICC curves being saturated, reducing the ability to detect item-level differentiation. Future research can build on this by involving a broader and more diverse participant pool to deepen understanding of the varied needs and experiences of individuals living with PsA. Particularly, including more testing with patients is necessary, which is essential to the proper design and effectiveness of the games despite the recruiting challenges associated with this task. Disparities in technology literacy among participants, especially health care professionals, also posed a challenge. Providing training or educational materials on the technologies used in the serious games could bridge this gap and improve feedback quality. Moreover, as the study is still at the prototype stage, further testing and clinical validation with patients with PsA through pilot studies and trials will be essential to refine the games based on real-world usability data. Finally, the proposed game designs may not fully consider the physical limitations of patients with PsA, such as joint stiffness and swelling, which can affect usability. Future iterations should incorporate adaptive game design features like customizable controls to improve accessibility.

Implications

The research on co-designing serious mobile games for patients with PsA presented here has several important implications across various domains.

From a managerial standpoint, the study underlines the importance of interdisciplinary collaboration in developing effective digital health interventions. Managers in health care and technology sectors should foster partnerships between clinicians, researchers, and game developers to leverage their combined expertise. This collaborative approach can lead to more innovative and user-centered solutions [47]. Additionally, the agile co-design methodology highlighted in the study can be adopted by managers to ensure continuous feedback and iterative improvements, enhancing the overall quality and relevance of the developed digital solutions [48]. By promoting a culture of collaboration and iterative development, managers can drive the successful implementation of digital health tools that meet the specific needs of patients [48,49].

From a financial perspective, the development and implementation of NoPain Games for patients with PsA could lead to significant cost savings in health care. By providing an engaging and effective tool for pain management and cognitive stimulation, these games can reduce the reliance on more expensive treatments and interventions [50]. Improved patient adherence to therapeutic activities can also lead to better health outcomes, potentially decreasing the frequency of hospital visits and the need for additional medical support. Furthermore, the scalability of mobile games means that they can be distributed widely at a relatively low cost, making them an economically viable option for health care providers and patients alike [51]. Investing in the development of such digital health solutions can yield long-term financial benefits for both health care systems and patients.

From the clinical point of view, while the NoPain Games presented here are still in the prototype stage and have not yet been used in clinical practice, their potential therapeutic benefits are promising. The proposed serious games are designed to provide pain distraction, cognitive stimulation, and stress reduction, which could significantly improve the quality of life for patients with PsA. Future clinical trials and real-world testing will be crucial to validate these benefits and refine the game mechanics based on patient feedback [52]. If proven effective, these serious games could be integrated into treatment plans as complementary therapy, offering a novel approach to managing chronic pain and enhancing cognitive function. This research also highlights the importance of adaptive difficulty levels and personalized gameplay experiences, which could be further explored in clinical settings to tailor the interventions to individual patient needs [53].

Finally, from a societal perspective, the emphasis on accessibility and inclusive design in the NoPain Games highlights the importance of creating digital health solutions that are usable by individuals with varying physical abilities. By addressing the specific needs of patients with PsA, the games promote greater equity in health care access and support for chronic pain management. This inclusive approach can serve as a model for developing other digital health interventions that address the needs of diverse patient populations [54]. Moreover, the widespread adoption of such games can raise awareness about the potential of serious games in health care, encouraging further innovation and investment in this field [51]. By making effective pain management tools accessible to a broader audience, the research can contribute to improved health outcomes and quality of life for many individuals living with chronic conditions.

In contrast to prior studies that focus on general chronic pain, this research is the first to co-design mobile serious games specifically for patients with PsA, integrating both expert and patient perspectives. It contributes to the field by demonstrating how agile, participatory design can yield tailored, accessible, and therapeutically meaningful digital interventions for a rheumatic population largely overlooked in serious game development.

Conclusions

The proposed NoPain Games hold significant potential for alleviating chronic pain in individuals with PsA by providing engaging and immersive gameplay experiences aimed at reducing discomfort. This study highlights the collaborative efforts of researchers and clinical/technical experts in designing NoPain Games as cocreators. It introduces 2 serious game prototypes developed using agile methodology and co-design principles, incorporating expert feedback. The study explores the cocreation process, presenting initial findings through storyboards, game visualizations, and prototypes informed by the collected input. Future work includes obtaining further patient feedback on the prototypes and conducting real-world testing to evaluate their feasibility, acceptability, and overall user satisfaction. The findings highlight the critical role of managing game difficulty, which can be addressed through the integration of a dynamic difficulty adjustment system [38,39] in developing the game to customize challenges based on each patient's condition. Moreover, potential additional features could include integrating smartwatch-based biometric data, such as stress levels and heart rate, to personalize gameplay experiences and provide valuable clinical insights.

Acknowledgments

The authors would like to thank and acknowledge all members of the iPROLEPSIS Consortium for their valuable contributions. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency. Neither the European Union nor the European Health and Digital Executive Agency can be held responsible for them. The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Funding

This work received funding from the European Union under the Horizon Europe Grant agreement number 101095697 (iPROLEPSIS: Psoriatic Arthritis Inflammation Explained through Multi-Source Data Analysis guiding a Novel Personalized Digital Care Ecosystem).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy restrictions and consortium policies on data sharing, but are available from the corresponding author on reasonable request.

Authors' Contributions

BR and SBD contributed to the study conceptualization and design; BR and SBD collected the data; BR, S Gomes, and SBD contributed to project administration. BR, S Gomes, MSV, FM, RGC, and SBD contributed to data curation and formal analysis; LH and SBD contributed to funding acquisition; BR, S Gomes, MSV, FM, RGC, LH, and SBD contributed to the methodology; BR, S Gomes, and SBD contributed to writing the original draft; BR, S Gomes, MSV, FM, RGC, S Gama, LH, and SBD contributed to reviewing and editing the paper; all authors reviewed and approved the final version of the manuscript.

We thank the collaborators from iPROLEPSIS Consortium: Leontios J. Hadjileontiadis, Vasileios Charisis, Georgios Apostolidis, Nikos Athanasopoulos, Eleni Vasileiou, Despina Petsini, Evdokimos Konstantinidis, Konstantina Tsimpita, Panagiotis Bamidis,

Jolanda Luime, Patty de Groot, Ilja Tchetverikov, Batoul Hojeij, Andreas Raptopoulos, Dimitrios Karamitros, Sofia Balula Dias, Samuel Gomes, Bárbara Ramalho, Marta Vicente, Filomena Carnide, Fátima Baptista, Kosmas Dimitropoulos, Dimitris Konstantinidis, Laura Coates, Cátia Gonçalves, Ana Rodrigues, Francesca Levi-Schaffer, Prince Ofori, Silvia Reis, Hugo Silva, Daniel Osório, Rodrigo Braga, Nikos Melanitis, Thanos Vidakis, Alex Bensenousi, Kristina Leipuviene, Ioannis Drivas, Amalia Ntemou, and Nikoleta Tsampanaki.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured guide script for the cocreation session.
[DOCX File, 2785 KB - games_v14i1e75072_app1.docx]

Multimedia Appendix 2
Online consent form.
[PNG File, 306 KB - games_v14i1e75072_app2.png]

References

1. Eichenberg C, Grabmayer G, Green N. Acceptance of serious games in psychotherapy: an inquiry into the stance of therapists and patients. Telemed J E Health 2016 Nov;22(11):945-951. [doi: 10.1089/tmj.2016.0001] [Medline: 27045452]
2. Brauner P, Ziefle M. Social acceptance of serious games for physical and cognitive training in older adults residing in ambient assisted living environments. J Public Health (Berl) 2022 Jan;30(1):63-75. [doi: 10.1007/s10389-021-01524-y]
3. Eichenberg C, Schott M. Serious games for psychotherapy: a systematic review. Games Health J 2017 Jun;6(3):127-135. [doi: 10.1089/g4h.2016.0068] [Medline: 28628385]
4. Neri SG, Cardoso JR, Cruz L, et al. Do virtual reality games improve mobility skills and balance measurements in community-dwelling older adults? systematic review and meta-analysis. Clin Rehabil 2017 Oct;31(10):1292-1304. [doi: 10.1177/0269215517694677] [Medline: 28933612]
5. Pacheco TBF, de Medeiros CSP, de Oliveira VHB, Vieira ER, de Cavalcanti FAC. Effectiveness of exergames for improving mobility and balance in older adults: a systematic review and meta-analysis. Syst Rev 2020 Jul 18;9(1):163. [doi: 10.1186/s13643-020-01421-7] [Medline: 32682439]
6. Gromala D, Tong X, Shaw C, Amin A, Ulas S, Ramsay G. Mobius Floe: an immersive virtual reality game for pain distraction. Electronic Imaging 2016;28:1-5. [doi: 10.2352/ISSN.2470-1173.2016.4.ERVR-413]
7. Saragih ID, Suarilah I, Saragih IS, Lin YK, Lin CJ. Efficacy of serious games for chronic pain management in older adults: a systematic review and meta-analysis. J Clin Nurs 2024 Mar;33(3):1185-1194. [doi: 10.1111/jocn.17012] [Medline: 38291564]
8. Vivekanantham A, McGagh D, Coates LC. Current treatments and recommendations for Psoriatic Arthritis. Best Pract Res Clin Rheumatol 2021 Jun;35(2):101680. [doi: 10.1016/j.berh.2021.101680]
9. Hetland ML. Psoriatic arthritis: still room for improvement. Lancet 2020 May;395(10235):1463-1465. [doi: 10.1016/S0140-6736(20)30800-X]
10. Fox S, Brown LJE, Antrobus S, et al. Co-design of a smartphone app for people living with dementia by applying agile, iterative co-design principles: development and usability study. JMIR Mhealth Uhealth 2022 Jan 14;10(1):e24483. [doi: 10.2196/24483] [Medline: 35029539]
11. Martin A, Caon M, Adorni F, et al. A mobile phone intervention to improve obesity-related health behaviors of adolescents across Europe: iterative co-design and feasibility study. JMIR Mhealth Uhealth 2020 Mar 2;8(3):e14118. [doi: 10.2196/14118] [Medline: 32130179]
12. Vasileiou E, Dias SB, Hadjidimitriou S, et al. Novel digital biomarkers for fine motor skills assessment in psoriatic arthritis: The DaktylAct touch-based serious game approach. IEEE J Biomed Health Inform 2025 Jan;29(1):128-141. [doi: 10.1109/JBHI.2024.3487785] [Medline: 39471112]
13. Gold JI, Belmont KA, Thomas DA. The neurobiology of virtual reality pain attenuation. Cyberpsychol Behav 2007 Aug;10(4):536-544. [doi: 10.1089/cpb.2007.9993]
14. Gunst M, De Meyere I, Willems H, Schoenmakers B. Effect of exergaming on wellbeing of residents in a nursing home: a single blinded intervention study. Aging Clin Exp Res 2022 Jan;34(1):151-157. [doi: 10.1007/s40520-021-01903-1] [Medline: 34156650]
15. Hoffman HG, Richards TL, Van Oostrom T, et al. The analgesic effects of opioids and immersive virtual reality distraction: evidence from subjective and functional brain imaging assessments. Anesth Analg 2007 Dec;105(6):1776-1783. [doi: 10.1213/01.ane.0000270205.45146.db] [Medline: 18042882]

16. Beltran-Alacreu H, Navarro-Fernández G, Godia-Lledó D, et al. A serious game for performing task-oriented cervical exercises among older adult patients with chronic neck pain: development, suitability, and crossover pilot study. JMIR Serious Games 2022 Feb 1;10(1):e31404. [doi: 10.2196/31404] [Medline: 35103608]
17. Peña J, Koebner I, Weisman W. Using digital art and attachment priming in a web-based serious game to reduce pain and social disconnection in individuals with chronic pain and loneliness: randomized controlled trial. JMIR Serious Games 2024 Nov 27;12:e52294. [doi: 10.2196/52294] [Medline: 39602564]
18. Maskeliūnas R, Damaševičius R, Kulikajevas A, Marley J, Larsson C. Evaluation of MyRelief serious game for better self-management of health behaviour strategies on chronic low-back pain. Informatics (MDPI) 2022;9(2):40. [doi: 10.3390/informatics9020040]
19. Ijaz K, Wang Y, Milne D, Calvo RA. Competitive vs affiliative design of immersive VR exergames. : Springer; 2016 Presented at: Serious Games: Second Joint International Conference, JCSG 2016; Sep 26-27, 2016 p. 140-150. [doi: 10.1007/978-3-319-45841-0_13]
20. Tuah NM, Ahmedy F, Gani A, Yong LN. A survey on gamification for health rehabilitation care: applications, opportunities, and open challenges. Information 2021;12(2):91. [doi: 10.3390/info12020091]
21. Tessarolo F, Petsani D, Conotter V, et al. Developing ambient assisted living technologies exploiting potential of user-centred co-creation and agile methodology: the CAPTAIN project experience. J Ambient Intell Human Comput 2024 May;15(5):2661-2676. [doi: 10.1007/s12652-021-03649-0]
22. iPROLEPSIS. URL: https://www.iprolepsis.eu [accessed 2025-03-09]
23. Hadjileontiadis LJ, Charisis V, Hadjidimitriou S, et al. European advances in digital rheumatology: explainable insights and personalized digital health tools for psoriatic arthritis. EClinicalMedicine 2025 Jun;84:103243. [doi: 10.1016/j.eclinm.2025.103243] [Medline: 40496877]
24. Ramalho B, Vicente M, Escobar H, et al. Developing exergames for psoriatic arthritis using agile storyboarding and game design processes. 2024 Jun 26 Presented at: PETRA ’24. [doi: 10.1145/3652037.3663939]
25. Ramalho B, Vicente M, Escobar H, et al. Developing sensorimotor art games for psoriatic arthritis using agile storyboarding and game co-design processes. 2024 Nov 13 Presented at: DSAI 2024. [doi: 10.1145/3696593.3696603]
26. Ramalho B, Vicente M, Escobar H, et al. “The kite” breathing serious game: agile co-design for psoriatic arthritis. 2024 Presented at: 2024 IEEE 22nd Mediterranean Electrotechnical Conference (MELECON. [doi: 10.1109/MELECON56669.2024.10608739]
27. Knapp, J, Zeratsky J, Kowitz B. Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days: Simon and Schuster; 2016.
28. Craft B, Cairns P. Sketching sketching: outlines of a collaborative design method. : BCS Learning and Development; 2009.
29. JW C, J DC. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches: SAGE; 2018.
30. Brooke J. SUS: a quick and dirty usability scale. : Digital Equipment Corporation; 1986.
31. IPROLEPSIS NoPain Games SUS questionnaire. Google Docs. URL: https://forms.gle/hXbaMKPeaRQEfAEJ9 [accessed 2026-01-12]
32. Reise S, Moore T. Item response theory. In: Cooper H, Coutanche M, McMullen L, Panter A, Rindskopf D, Sher K, editors. Handbook of Research Methods in Psychology: Foundations, Planning, Measures, and Psychometrics: American Psychological Association; 2023:809-835. [doi: 10.1037/0000318-037]
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282. [doi: 10.11613/BM.2012.031] [Medline: 23092060]
35. Birnbaum A. Some latent trait models and their use in inferring an examinee’s ability. In: Lord F, Novick M, editors. Statistical Theories of Mental Test Scores 1968:397-479.
36. 40 years of Simon, the electronic game that never stops reinventing itself. Fast Compnay. URL: https://www.fastcompany.com/40491956 [accessed 2025-02-10]
37. Abd-Alrazaq A, Alhuwail D, Al-Jafar E, et al. The effectiveness of serious games in improving memory among older adults with cognitive impairment: systematic review and meta-analysis. JMIR Serious Games 2022 Aug 9;10(3):e35202. [doi: 10.2196/35202] [Medline: 35943792]
38. Lopes JC, Lopes RP. A review of dynamic difficulty adjustment methods for serious games. Presented at: Optimization, Learning Algorithms and Applications; Oct 24-25, 2022. [doi: 10.1007/978-3-031-23236-7_11]
39. Gobel S, Hardy S, Wendel V, Mehm F, Steinmetz R. Serious games for health: personalized exergames. Presented at: MM ’10: ACM Multimedia Conference; Oct 25-29, 2010. [doi: 10.1145/1873951.1874316]
40. Nunes F, Rato Grego P, Araújo R, Silva PA. Self-report user interfaces for patients with rheumatic and musculoskeletal diseases: app review and usability experiments with mobile user interface components. Comput Graph 2023 Dec;117:61-72. [doi: 10.1016/j.cag.2023.10.009]
41. Armstrong-Wood R, Messiou C, Kite A, et al. Smartphone accessibility: understanding the lived experience of users with cervical spinal cord injuries. Disabil Rehabil Assist Technol 2024 May;19(4):1434-1445. [doi: 10.1080/17483107.2023.2192246] [Medline: 37010939]

42. Hadjileontiadou S, Dias SB, Hadjileontiadis L. A conceptual framework (2D-ME) for explaining self-first and self-third person views of prototyping dynamics in serious games design: experimental case study. JMIR Serious Games 2023 Apr 24;11:e41824. [doi: 10.2196/41824] [Medline: 37093627]
43. Hodent C. The Gamer's Brain: How Neuroscience and UX Can Impact Video Game Design: CRC Press; 2017. [doi: 10.1201/9781315154725]
44. Smith ET, Basak C. A game-factors approach to cognitive benefits from video-game training: a meta-analysis. PLOS ONE 2023;18(8):e0285925. [doi: 10.1371/journal.pone.0285925] [Medline: 37531408]
45. Csikszentmihalyi and M, Csikszentmihalyi IS. Optimal experience: psychological studies of flow in consciousness. : Cambridge University Press; 1992.
46. Pinge A, Gad V, Jaisighani D, Ghosh S, Sen S. Detection and monitoring of stress using wearables: a systematic review. Front Comput Sci 2024;6. [doi: 10.3389/fcomp.2024.1478851]
47. Krause-Jüttler G, Weitz J, Bork U. Interdisciplinary collaborations in digital health research: mixed methods case study. JMIR Hum Factors 2022 May 4;9(2):e36579. [doi: 10.2196/36579] [Medline: 35507400]
48. Kilfoy A, Hsu TCC, Stockton-Powdrell C, Whelan P, Chu CH, Jibb L. An umbrella review on how digital health intervention co-design is conducted and described. NPJ Digit Med 2024 Dec 23;7(1):374. [doi: 10.1038/s41746-024-01385-1] [Medline: 39715947]
49. Oham C, Ejike OG. Creativity and collaboration in creative industries: proposing a conceptual model for enhanced team dynamics. Magna Sci Adv Res Rev 2024;12(1):185-194. [doi: 10.30574/msarr.2024.12.1.0172]
50. Vugts MA, Zedlitz AM, Joosen MC, Vrijhoef HJ. Serious gaming during multidisciplinary rehabilitation for patients with chronic pain or fatigue symptoms: mixed methods design of a realist process evaluation. J Med Internet Res 2020 Mar 9;22(3):e14766. [doi: 10.2196/14766] [Medline: 32149720]
51. Damaševičius R, Maskeliūnas R, Blažauskas T. Serious games and gamification in healthcare: a meta-review. Information 2023;14(2):105. [doi: 10.3390/info14020105]
52. Lane K, Majkowski R, Gruber J, et al. Using gamification to enhance clinical trial start-up activities. J Clin Trans Sci 2022;6(1):e75. [doi: 10.1017/cts.2022.405]
53. Fisher N, Kulshreshth AK. Exploring dynamic difficulty adjustment methods for video games. Virtual Worlds ;3(2):230-255. [doi: 10.3390/virtualworlds3020012]
54. Waller S, Bradley M, Hosking I, Clarkson PJ. Making the case for inclusive design. Appl Ergon 2015 Jan;46:297-303. [doi: 10.1016/j.apergo.2013.03.012] [Medline: 23538129]

Abbreviations

AI: artificial intelligence
AI-PGS: Artificial Intelligence-Personalized Game Suite
ICC: item characteristic curve
PsA: psoriatic arthritis
SUS: System Usability Scale
UI: user interface
VR: virtual reality

Edited by S Brini; submitted 28.Mar.2025; peer-reviewed by E Benitez-Guerrero, T Adeel; accepted 11.Nov.2025; published 30.Jan.2026.

Please cite as:
Ramalho B, Gomes S, Silva Vicente M, Magalhães F, Gonçalves Costa R, Gama S, Charisis V, Hadjileontiadis L, B Dias S, iPROLEPSIS Consortium
Co-Designing Mobile Serious Games to Support Patients With Psoriatic Arthritis and Chronic Pain: Mixed Methods Study
JMIR Serious Games 2026;14:e75072
URL: https://games.jmir.org/2026/1/e75072
doi:10.2196/75072

Original Paper

Efficacy and Safety of a Video Game–Like Digital Therapy Intervention for Chinese Children With Attention-Deficit/Hyperactivity Disorder: Single-Arm, Open-Label Pre-Post Study

Mengyi Bao[1*], MM; Shang Feng[2*], PhD; Jiaheng Wang[3], PhD; Jia Ye[4], MM; Jiangping Wang[1], MD; Wenyu Li[2], BEng; Kewen Jiang[1], MD; Lin Yao[3,5,6], PhD

[1]Children's Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

[2]SDO Digital Therapeutics Innovation Center, Shanghai, Shanghai, China

[3]the MOE Frontiers Science Center for Brain and Brain-Machine Integration, the College of Computer Science, and the College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, Zhejiang, China

[4]Shengzhou People's Hospital (the First Affiliated Hospital of Zhejiang University Shengzhou Branch), Shengzhou, Zhejiang, China

[5]Nanhu Brain-Computer Interface Institute, Hangzhou, Zhejiang, China

[6]Affiliated Mental Health Center and Hangzhou Seventh People's Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

[*]these authors contributed equally

Corresponding Author:
Lin Yao, PhD
the MOE Frontiers Science Center for Brain and Brain-Machine Integration, the College of Computer Science, and the College of Biomedical Engineering and Instrument Science
Zhejiang University
No. 866, Yuhangtang Road, Xihu District (Post code 310000)
Hangzhou, Zhejiang,
China
Phone: 86 0571 87952838
Email: lin.yao@zju.edu.cn

Abstract

Background: The digital therapy of attention-deficit/hyperactivity disorder (ADHD) based on a "self-adaptive multitasking training paradigm" has been developed to improve the cognitive functional impairments and attention deficits of children with ADHD. However, the efficacy and safety of such treatment for Chinese patients remain untested.

Objective: This study aimed to preliminarily evaluate the actual intervention effects of a video game–like training software (ADHD-DTx) for children with ADHD aged 6-12 years as the first nationally certified digital therapeutics medical device for ADHD in China. We performed a single-arm, open-label efficacy and safety study.

Methods: This is a single-arm, open-label, pre-post efficacy and safety study. A total of 97 participants were included in the analysis. Participants received digital therapy (ADHD-DTx) and basic behavioral parent training for 4 weeks (25 min/day, ≥5 times/week) without medication. The efficacy outcomes included the Test of Variables of Attention (TOVA), Swanson, Nolan, and Pelham Questionnaire, version 4 (SNAP-IV), Weiss Functional Impairment Rating Scale (WFIRS), and Conner's Parent Symptom Questionnaire (PSQ). Safety-related events were monitored during and after the trial.

Results: From day 0 (baseline) to day 28, the population TOVA Attention Performance Index exhibited statistically significant improvement (from mean –4.15, SE of the mean [SEM] 0.32 to mean –1.70, SEM 0.30; t_{94}=–8.78; n=95; P<.001); the population total, inattention (AD), hyperactivity/impulsivity (HD), and oppositional defiant disorder (ODD) scores of SNAP-IV all significantly improved (total: from mean 1.33, SEM 0.05 to mean 1.09, SEM 0.05; t_{96}=5.32; P<.001; AD: from mean 1.71, SEM 0.06 to mean 1.44, SEM 0.06; t_{96}=4.44; P<.001; HD: from mean 1.38, SEM 0.07 to mean 1.05, SEM 0.06; t_{96}=5.96; P<.001; ODD: mean 0.84, SEM 0.05 to mean 0.75, SEM 0.05; Z=2.47; P=.03; n=97); for WFIRS results, domains of "family" and "social activities" showed significant population improvement (family: from mean 0.75, SEM 0.05 to mean 0.65, SEM 0.04; Z=2.80; P=.01; social activities: from mean 0.56, SEM 0.05 to mean 0.45, SEM 0.05; Z=2.91; P=.01; n=97); for PSQ results, domains of "learning problem," "psychosomatic problem," "impulsivity-hyperactivity," and "hyperactivity index" showed significant improvement (learning problem: from mean 1.72, SEM 0.06 to mean 1.57, SEM 0.06; Z=2.42; P=.03; psychosomatic problem: from mean 0.40, SEM 0.03 to mean 0.32, SEM 0.03; Z=2.66; P=.02; impulsivity-hyperactivity: from mean 0.94, SEM 0.06 to mean 0.80, SEM 0.06; Z=2.49; P=.03; hyperactivity index: from mean 1.06, SEM 0.05 to mean 0.92, SEM 0.05; Z=2.90; P=.01; n=97). No device-related adverse event or severe adverse event was observed or reported during or after the intervention.

Conclusions: This study preliminarily suggested the significant improvements of ADHD symptoms and attention function after 4 weeks of ADHD-DTx digital therapy combining basic behavioral parent training with satisfying safety outcomes.

(JMIR Serious Games 2026;14:e76114) doi:10.2196/76114

KEYWORDS

ADHD; attention; attention-deficit/hyperactivity disorder; cognitive training; digital therapy; rehabilitation

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that commonly occurs in childhood and is characterized by inattention, hyperactivity, and impulsivity. The primary treatment for ADHD is pharmacotherapy, using methylphenidate, dexamphetamine, or atomoxetine as the first-line drugs [1]. However, pharmacotherapy for ADHD could cause adverse drug reactions such as digestive system issues, irritability, palpitations, and headaches, and more importantly, pharmacotherapy does not target the core functional deficits of patients with ADHD [2]. A recent study also suggested that beginning treatment with behavioral intervention may produce better outcomes overall than beginning treatment with medication [3]. Therefore, many patients with ADHD and their parents are looking forward to alternative interventions [4] like behavior therapy, neurofeedback, counseling, and, more recently, digital therapy [5].

The digital therapy of ADHD has been developed to improve the cognitive functional impairments and attention-control deficits of children with ADHD. In 2020, the US Food and Drug Administration (FDA) approved EndeavorRx (AKL-T01), the first video game–like training software for children with ADHD aged 8-12 years. Growing evidence has suggested that digital therapy could provide a safe and effective intervention to improve functional performance (such as attention, working memory, etc) and problematic behavior of children with ADHD, with minimal risk of adverse events (AEs) compared to pharmacotherapy [6-9]. In 2023, the Chinese National Medical Products Administration (NMPA) approved the "Attention Enhancement Training Software" (ADHD-DTx), a video game–like training software for children with ADHD aged 6-12 years, as the first nationally certified digital therapeutics medical device for ADHD in China.

The neurophysiological mechanisms of EndeavorRx and ADHD-DTx are both based on the self-adaptive multitasking training paradigm (NeuroRacer), which consists of 2 tasks: the "driving task" (sustained attention) and the go/no-go "sign task" (signal detection, attention-shifting, and inhibition control). Multiple studies have revealed that this training paradigm could improve participants' cognitive control abilities (enhanced sustained attention and working memory) after 1-month training [6-8,10], as well as increase the frontal midline theta (FM) power, the neuromarker of sustained attention and cognitive control [9,10]. Therefore, this paradigm has been used for the intervention of ADHD in a video game–like form to improve compliance in children. Besides the "driving task" and "sign task," ADHD-DTx also included a third task: the "digit cancellation task" (a widely used attention assessment and training method in clinical practice) [11] to further enhance the training effect of attention function.

To preliminarily evaluate the actual intervention effects of ADHD-DTx in Chinese children with ADHD aged 6-12 years, we performed a single-arm, open-label efficacy and safety study in the Children's Hospital of Zhejiang University School of Medicine (Hangzhou, China) in 2021. This study provided valuable efficacy and safety data of ADHD-DTx (and ADHD digital therapy based on the self-adaptive multitasking training paradigm) for the first time in Chinese children with ADHD. The efficacy data suggested by this pilot study provided a critical contribution to the design of the following randomized, double-blinded, parallel-controlled clinical trials, which were conducted during 2022 and 2023 (Feng S, PhD, unpublished data, 2025). The Chinese NMPA approval of ADHD-DTx (as the first Chinese ADHD digital therapy medical device) was based on the results of a key Good Clinical Practice clinical trial conducted during 2022-2023 (Feng S, PhD, unpublished data, 2025).

All participants had to be off any ADHD medication and without other significant comorbid psychiatric diagnoses. Included participants were treated with ADHD-DTx therapy and basic behavioral parent training (BPT; positive reinforcement training at home, required according to the ethical consideration of the regulatory agency) [12,13]. Efficacy outcomes included the computerized attention test: Test of Variables of Attention (TOVA) [14,15] and classic scales: (1) the Swanson, Nolan, and Pelham Questionnaire, version 4 (SNAP-IV) [16-18]; (2) the Weiss Functional Impairment Rating Scale (WFIRS) [19]; and (3) the Conner's Parent Symptom Questionnaire (PSQ) [20,21]. Safety outcomes included the proportions of device-related AEs or severe AEs.

Methods

Overview

This study was a single-arm, open-label study in children (aged 6-12 years) with a confirmed diagnosis of ADHD (as per the DSM-5 [*Diagnostic and Statistical Manual of Mental Disorders* {Fifth Edition}) in the Children's Hospital of Zhejiang

University School of Medicine (Hangzhou, China) in 2021. Participants had to be off any ADHD medication and not present other significant comorbid psychiatric diagnoses. All participants had an IQ score of ≥80 (per the Wechsler Intelligence Test) and an Attention-Deficit/Hyperactivity Disorder Rating Scale IV (ADHD-RS-IV) total score of >28.

Study Design

The study planned to enroll about 100 participants with a confirmed diagnosis of ADHD. Participants received digital therapy (ADHD-DTx, a video game–like training software running on an Android tablet) and basic BPT (positive reinforcement training at home) for 4 weeks (25 min/day, ≥5 times/week). Efficacy outcomes were measured on day 0 (baseline visit) and day 28 (after 4-week treatment). Safety-related events were monitored during and after the trial.

Participants

Eligible patients were male or female children aged 6-12 years with a confirmed diagnosis of ADHD (as per the DSM-V). Participants had to be off any ADHD medication (for at least 4 weeks before the baseline visit) and not present other significant comorbid psychiatric diagnoses. All participants had an IQ score of ≥80 (per the Wechsler Intelligence Test) and an ADHD-RS-IV total score of ≥28. Complete inclusion and exclusion criteria are in Multimedia Appendix 1.

Procedures

The ADHD-DTx intervention was preinstalled on Huawei MatePad tablets (Huawei). At the baseline visit, eligible patients were instructed to use ADHD-DTx for about 10 minutes while a study coordinator monitored the session to ensure that patients could follow the rules of ADHD-DTx. Patients were further assessed by ADHD-related scales (symptoms and impairments) and TOVA (attention functioning) at the baseline visit.

Afterward, patients received the ADHD-DTx digital treatment and basic BPT at home for 4 weeks (25 min/day, ≥5 times/week). Basic BPT was required according to the ethical considerations of the regulatory agency. The daily training task consisted of five 4- to 5-minute multitasking missions (total time on task was about 25 minutes). Compliance was monitored remotely using the network by investigators, and daily training reminders were sent to patients' caregivers manually or automatically. In accordance with a strict, predefined protocol, any participant who failed to complete the required training (25 min/day, ≥5 times/week) and persistently ignored reminders was classified as noncompliant and withdrawn from the trial.

ADHD-DTx is a digital therapeutic that uses a proprietary algorithm designed to improve attention control by training interference management (multitasking). ADHD-DTx mechanisms have been described previously. In brief, users multitask by responding to a perceptual discrimination targeting task and a simultaneous sensory motor navigation task. Users advance by reducing interference costs (closing the performance gap between multitasking and single-tasking), and real-time and periodic recalibration occurs to maintain an optimal difficulty level.

BPT consisted of parent-training courses (focusing on teaching parents specialized child management techniques primarily involving contingency management, such as behavior management principles, parental attending skills and home token system), aerobic exercises (such as jogging, swimming, or rope-skipping 40 min/day, ≥4 times/week), listening-retelling training (the child was asked to accurately retell sentences presented orally by parent, and the sentences were gradually made longer and more detailed to train the child's auditory attention and memory span), and reading-aloud training (the child was asked to read texts aloud to maintain focus and reduce mind-wandering, thereby promoting the child's sustained attention) [22-24]. BPT is a regular basic treatment for all children with ADHD in clinical practice.

Safety-related events such as vomiting, dizziness, headache, palpitation, addiction, frustration, eye discomfort, and other similar symptoms were monitored during and after the trial, as described: (1) during the 4-week intervention period, all participants' parents were contacted daily by a research assistant using online social software to collect safety-related events; and (2) after the intervention period, all participants' parents were contacted monthly by a research assistant using the telephone for 3 months to collect long-term safety-related events.

Outcomes

To effectively evaluate the impact of treatment on the core symptoms, problematic behaviors, and functional deficits of children with ADHD, and also ensure the reliability, validity, and objectivity of data, we chose both the computerized attention test and classic scales as the efficacy outcomes.

The primary end point was the improvement of attentional functioning as measured by TOVA from baseline to day 28. TOVA is a computerized, objective test of attentional functioning and has been globally used in clinical and academic institutions [22-24]. The outcomes of TOVA could objectively reflect the functional training effect of ADHD-DTx and provide essential information about the efficacy of the intervention.

Secondary end points included the improvements of classic scales that assess ADHD-related symptoms, functional impairments, and problematic behaviors from baseline to day 28. The SNAP-IV is widely used as the key assessment of ADHD core symptoms, including 3 subsets: (1) inattention (AD), (2) hyperactivity/impulsivity (HD), and (3) oppositional defiant disorder (ODD) [16,17]. WFIRS is a multidimensional, ADHD-specific, functional impairment assessment scale, including 6 domains: family, school, life skills, child's self-concept, social activities, and risky activities [19]. The PSQ has been widely used to assess problematic behaviors related to ADHD, including 6 domains: conduct problem, learning problem, psychosomatic problem, impulsivity-hyperactivity, anxiety, and hyperactivity index [20,21]. The selection of classic, widely validated scales could provide rich information about the behavioral symptoms, and importantly, the outcomes of these scales could reflect the influence of daily life by the training of ADHD-DTx, which is crucial for the rehabilitation of children with ADHD. More detailed descriptions of the outcome measurements can be found in Multimedia Appendix 2.

Statistical Analysis

All analyses were performed according to a prespecified statistical analysis plan. Unless otherwise indicated, statistical comparisons used a 2-tailed significance test evaluated at the 95% level of confidence. All analyses were conducted using a complete case analysis. In no situation were missing data to be imputed. Student *t* test was performed only if the analyzed data passed the normality test (Shapiro-Wilk test); otherwise, the Wilcoxon nonparametric test would be used. The chi-square test was used for the statistical inference for counting data. Multiple comparisons corrections (using the false discovery rate method) were performed for all the secondary efficacy end points [25].

The primary efficacy end point for each participant was the change in the TOVA Attention Performance Index (API) from baseline to day 28, defined as the score on day 28 minus the score at baseline. Missing data were not imputed. Participants with missing data either on day 0 or day 28 would be excluded from the paired significance analysis. Unless otherwise indicated, the results of the efficacy analysis were summarized as mean (SE of the mean [SEM]). Significance was assessed with a 2-sided paired *t* test evaluated at the 95% level of confidence.

The following secondary efficacy end points were tested using the same technique outlined for the primary efficacy analysis: (1) change in SNAP-IV from baseline to day 28, (2) change in WFIRS from baseline to day 28, and (3) change in PSQ from baseline to day 28.

Ethical Considerations

The study was conducted in accordance with the International Conference on Harmonisation Regulations and was approved by the Institutional Review Board of Children's Hospital of Zhejiang University School of Medicine. All participants and their caregivers provided written informed consent prior to any study activities being conducted. The data of all participants were anonymous throughout the study. All participants received compensation for the examination fees. The study was subject to independent supervision by regulatory agencies (the Institutional Review Board and the Office of Clinical Trial Institution of Children's Hospital of Zhejiang University School of Medicine) throughout the entire process. Authors who were employed by SDO Digital Therapeutics (the developer of ADHD-DTx) did not participate in any data collection process, and the data analysis results were checked and confirmed by authors from the Children's Hospital of Zhejiang University School of Medicine. To safeguard the design details and clinical parameters of the investigational product (ADHD-DTx) and mitigate potential commercial risks, the producer requested a delay in the clinical trial registration until after market approval was obtained from the NMPA of China. Following the product's market approval in 2023, the trial was submitted for registration with the Chinese Clinical Trial Registry, where it is currently pending review.

Results

Participants

A total of 114 participants were screened for inclusion in this study, with 110 meeting eligibility criteria (Figure 1). The sample size was determined comprehensively, referring to previous research using similar digital therapy for children with ADHD [8], the power of statistics, and the amount of available resources. The mean age of included participants was 7.78 (SD 1.14) years, and 90% (99/110) were male. The vast majority (106/110, 96.36%) of included participants were of the Han ethnic group, and other involved ethnic groups included She (2/110, 1.82%), Korean (1/110, 0.91%), and Tujia (1/110, 0.91%). The educational status of participants' parents: postgraduate (6/110, 5.45%), graduate or junior college (62/110, 56.36%), high school or vocational school (28/110, 25.45%), middle school (12/110, 10.91%), primary school, and below (2/110, 1.82%). Demographic characteristics of included participants are listed in Table 1.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the single-arm, open-label study of children with attention-deficit/hyperactivity disorder in China during 2021. TOVA: Test of Variables of Attention.

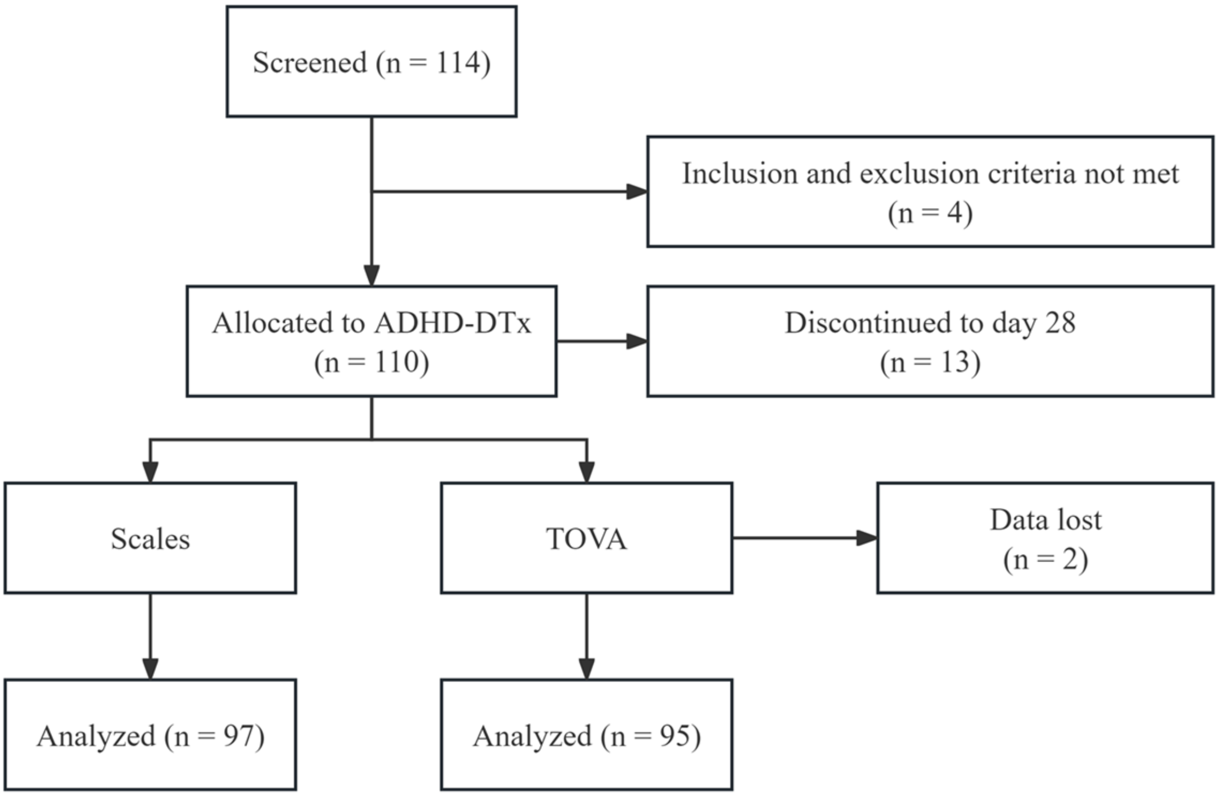


Table 1. Demographic characteristics of included participants.

Demographic	Value (n=110)
Age (years), mean (SD)	7.78 (1.14)
Gender, n (%)	
Male	99 (90)
Female	11 (10)
Ethnicity, n (%)	
Han	106 (96.36)
She	2 (1.82)
Korean	1 (0.91)
Tujia	1 (0.91)
Parental education level, n (%)	
Postgraduate	6 (5.45)
Graduate or junior college	62 (56.36)
High school or vocational school	28 (25.45)
Middle school	12 (10.91)
Primary school and below	2 (1.82)

Overall, 97/110 (88%) participants completed the study, 13 participants did not complete the study (lost to follow-up or noncompliance), and the TOVA data of 2 participants were lost due to device failure. All 97 participants who completed the study adhered strictly to the training protocol (25 min/day, ≥5 times/week for 4 weeks), and the average number of actual training days completed per participant was 20. Therefore, we finally collected and analyzed 97 participants' scales data, as well as 95 participants' TOVA data. All participants received

ADHD-DTx digital therapy (combined with basic BPT) in 2021 and remained off medication during the treatment period.

Efficacy Outcomes

In this study, efficacy outcomes were measured using the computerized attention test (TOVA) and classic scales: SNAP-IV, WFIRS, and PSQ. After a 4-week intervention, we found that the treatment significantly improved ADHD-related symptoms and impairments (all detailed statistical results are present in Table 2).

Table 2. Summary of efficacy outcomes during the 4-week treatment.

Efficacy outcome	Participants, n	Day 0, mean (SEM)	Day 28, mean (SEM)	*P* value	Test statistic
TOVA[a]-API[b]	95	–4.15 (0.32)	–1.70 (0.30)	<.001	–8.78 (94)[c]
Younger group	69	–4.15 (0.35)	–1.59 (0.35)	<.001	–7.90 (68)[c]
Older group	26	–4.15 (0.73)	–2.00 (0.59)	.001	–3.90 (25)[c]
Male group	84	–4.18 (0.34)	–1.74 (0.31)	<.001	–8.11 (83)[c]
Female group	11	–3.89 (1.08)	–1.38 (1.19)	.008	–3.31 (10)[c]
SNAP-IV[d] total	97	1.33 (0.05)	1.09 (0.05)	<.001	5.32 (96)[c]
Younger group	70	1.30 (0.05)	1.09 (0.06)	<.001	4.44 (69)[c]
Older group	27	1.39 (0.11)	1.11 (0.09)	.007	2.91 (26)[c]
Male group	86	1.35 (0.05)	1.12 (0.05)	<.001	4.78 (85)[c]
Female group	11	1.14 (0.08)	0.87 (0.16)	.03	2.59 (10)[c]
SNAP-IV AD[e]	97	1.71 (0.06)	1.44 (0.06)	<.001	4.44 (96)[c]
Younger group	70	1.72 (0.07)	1.42 (0.07)	<.001	4.77 (69)[c]
Older group	27	1.68 (0.11)	1.51 (0.12)	.23	1.22 (26)[c]
Male group	86	1.73 (0.06)	1.48 (0.07)	<.001	3.97 (85)[c]
Female group	11	1.57 (0.11)	1.20 (0.19)	.07	2.04 (10)[c]
SNAP-IV HD[f]	97	1.38 (0.07)	1.05 (0.06)	<.001	5.96 (96)[c]
Younger group	70	1.34 (0.07)	1.07 (0.08)	<.001	4.53 (69)[c]
Older group	27	1.49 (0.14)	1.01 (0.09)	<.001	3.98 (26)[c]
Male group	86	1.42 (0.07)	1.07 (0.06)	<.001	5.69 (85)[c]
Female group	11	1.09 (0.19)	0.89 (0.25)	.09	1.91 (10)[c]
SNAP-IV ODD[g]	97	0.84 (0.05)	0.75 (0.05)	.03	2.47[h]
Younger group	70	0.80 (0.06)	0.74 (0.06)	.17	1.38 (69)[c]
Older group	27	0.95 (0.13)	0.76 (0.11)	.01	2.50 (26)[c]
Male group	86	0.86 (0.06)	0.78 (0.06)	.06	1.88 (85)[c]
Female group	11	0.70 (0.10)	0.47 (0.08)	.04	2.02 (10)[c]
WFIRS[i]					
Family	97	0.75 (0.05)	0.65 (0.04)	.01	2.80[h]
School	97	0.78 (0.05)	0.72 (0.04)	.14	1.55[h]
Life skills	97	0.72 (0.04)	0.65 (0.04)	.09	1.85[h]
Child’s self-concept	97	0.65 (0.06)	0.59 (0.05)	.34	0.96[h]
Social activities	97	0.56 (0.05)	0.45 (0.05)	.01	2.91[h]
Risky activities	97	0.27 (0.03)	0.23 (0.03)	.14	1.54[h]
PSQ[j]					
Conduct problem	97	0.82 (0.05)	0.74 (0.05)	.06	2.04[h]
Learning problem	97	1.72 (0.06)	1.57 (0.06)	.03	2.42[h]
Psychosomatic problem	97	0.40 (0.03)	0.32 (0.03)	.02	2.66[h]

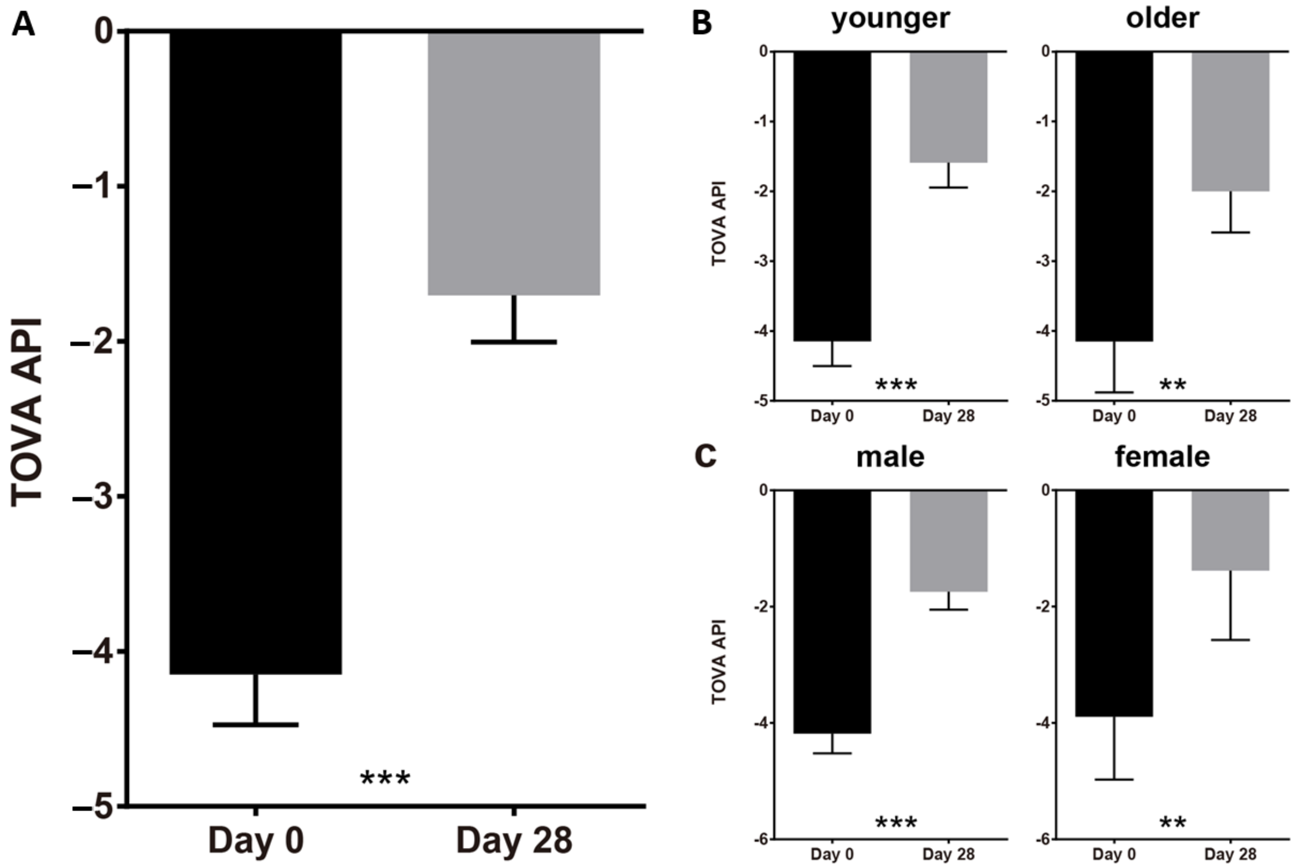
Efficacy outcome	Participants, n	Day 0, mean (SEM)	Day 28, mean (SEM)	*P* value	Test statistic
Anxiety	97	0.36 (0.04)	0.34 (0.04)	.17	1.40[h]
Impulsivity-hyperactivity	97	0.94 (0.06)	0.80 (0.06)	.03	2.49[h]
Hyperactivity index	97	1.06 (0.05)	0.92 (0.05)	.01	2.90[h]

[a]TOVA: Test of Variables of Attention.
[b]API: Attention Performance Index.
[c]*t* test (*df*).
[d]SNAP-IV: Swanson, Nolan, and Pelham Questionnaire, version 4.
[e]AD: inattention subscale of SNAP-IV.
[f]HD: hyperactivity/impulsivity subscale of SNAP-IV.
[g]ODD: oppositional defiant disorder subscale of SNAP-IV.
[h]*Z* value of Wilcoxon nonparametric test.
[i]WFIRS: Weiss Functional Impairment Rating Scale.
[j]PSQ: Conner's Parent Symptom Questionnaire.

Primary End Point: TOVA-API

We chose TOVA (an objective measurement of attention function) as the primary measurement of intervention efficacy, and the TOVA-API was chosen as the primary end point of this study (Figure 2).

Figure 2. Statistical results of the Test of Variables of Attention (TOVA) suggested significant improvement after 4 weeks of intervention. (A) TOVA Attention Performance Index (API) outcomes from day 0 to day 28; (B) TOVA-API of younger (aged 6-8 years) and older (aged 9-12 years) groups from day 0 to day 28; (C) TOVA-API of male and female groups from day 0 to day 28. Error bars indicate SEM. ***P*<.01, ****P*<.001.



The baseline data of TOVA-API passed the normality test (n=95; *P*=.36). From day 0 (baseline) to day 28, the population TOVA-API exhibited statistically significant improvement (Figure 2A, from mean –4.15, SEM 0.32 to mean –1.70, SEM 0.30; T=–8.78; n=95; *P*<.001), suggesting the efficacy of objective functional improvement of attention.

We conducted subgroup analysis of TOVA-API on age (Figure 2B; younger: 6-8 years; older: 9-12 years) and gender (Figure 2C; male and female) to investigate more detailed efficacy characteristics. We found that from day 0 (baseline) to day 28, both the younger and older groups exhibited statistically significant improvement (younger group: from mean –4.15, SEM 0.35 to mean –1.59, SEM 0.35; T=–7.90; n=69; P<.001; older group: from mean –4.15, SEM 0.73 to mean –2.00, SEM 0.59; T=–3.90; n=26; P=.001). The improvement of the younger group (pre-post difference=2.56) slightly exceeded that of the older group (pre-post difference=2.15), but no statistical difference was found (P=.53). As for the gender analysis, from day 0 (baseline) to day 28, both the male and female groups exhibited statistically significant improvement (male group: from mean –4.18, SEM 0.34 to mean –1.74, SEM 0.31; T=–8.11; n=84; P<.001; female group: from mean –3.89, SEM 1.08 to mean –1.38, SEM 1.19; T=–3.31; n=11; P=.008). The improvements of both groups showed no statistical difference (P=.93). We also analyzed the potential influence of parental education level (college degree or above: n=56; below college level: n=39) on TOVA-API and found no statistically significant difference (P=.78).

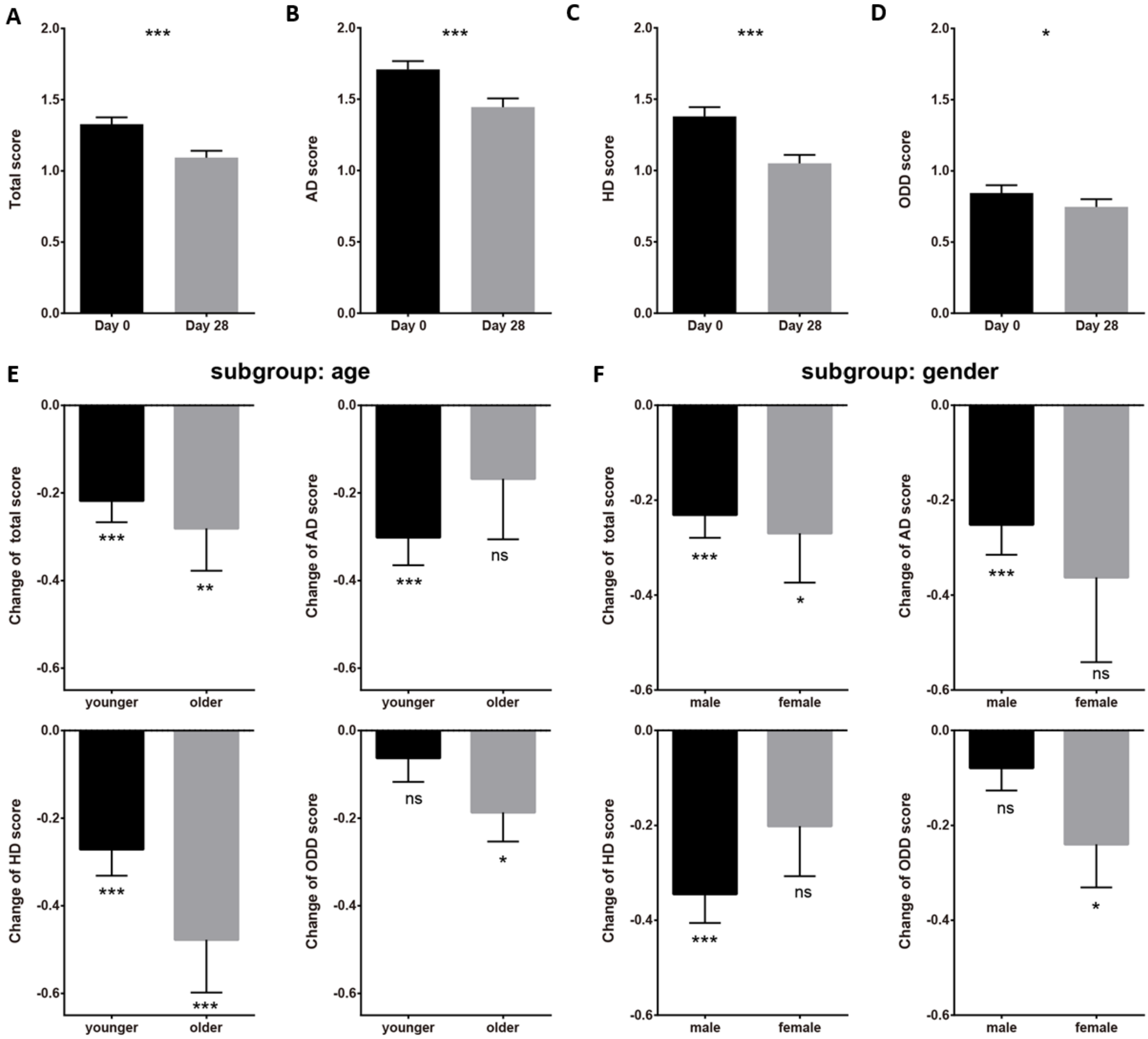
The above findings suggested that the 4-week intervention could significantly improve the attention function of children with ADHD (regardless of age or gender), as measured by TOVA-API.

Secondary End Points: SNAP-IV, WFIRS, and PSQ

We chose the results of classic clinical scales (SNAP-IV, WFIRS, and PSQ) as the secondary end points. Multiple comparisons corrections were performed for the statistical analysis results of the scale scores (see Methods section).

Using the SNAP-IV scale, we measured the treatment effect of ADHD core symptoms (Figure 3). The baseline data of the SNAP-IV total score, AD subscale, and HD subscale passed the normality test (n=97; P_{total}=.19; P_{AD}=.09; P_{HD}=.25), while that of the ODD subscale did not (n=97; P_{ODD}<.001). Therefore, we used the Student t test for total score, AD, and HD data, and the Wilcoxon nonparametric test for ODD data.

Figure 3. Statistical results of the Swanson, Nolan, and Pelham Questionnaire, version 4 (SNAP-IV), suggested significant improvement after 4 weeks of intervention. (A) SNAP-IV total score from day 0 to day 28; (B) SNAP-IV inattention (AD) score from day 0 to day 28; (C) SNAP-IV hyperactivity/impulsivity (HD) score from day 0 to day 28. (D) SNAP-IV oppositional defiant disorder (ODD) score from day 0 to day 28. (E) Subgroup analysis of age (younger: 6-8 years; older: 9-12 years) of SNAP-IV scores from day 0 to day 28. (F) Subgroup analysis of gender (male and female) of SNAP-IV scores from day 0 to day 28. Error bars indicate SEM. ns: nonsignificant. *P<.05, **P<.01, ***P<.001.



From day 0 to day 28, the population total score of SNAP-IV significantly improved (descended) from mean 1.33, SEM 0.05 to mean 1.09, SEM 0.05 (Figure 3A; T=5.32; n=97; *P*<.001); the population AD score of SNAP-IV significantly improved (descended) from mean 1.71, SEM 0.06 to mean 1.44, SEM 0.06 (Figure 3B; T=4.44; n=97; *P*<.001); the population HD score of SNAP-IV significantly improved (descended) from mean 1.38, SEM 0.07 to mean 1.05, SEM 0.06 (Figure 3C; T=5.96; n=97; *P*<.001); the population ODD score of SNAP-IV significantly improved (descended) from mean 0.84, SEM 0.05 to mean 0.75, SEM 0.05 (Figure 3D; *Z*=2.47; n=97; *P*=.03). The above results suggested that the treatment could significantly improve the core symptoms (attention deficit and hyperactivity) of children with ADHD.

We also conducted subgroup analysis of SNAP-IV on age (Figure 3E; younger: 6-8 years; older: 9-12 years) and gender (Figure 3F; male and female). It was notable that due to the small sample size of the female group (n=11) and older group (n=27), the efficacy of statistical inference might be limited.

For the SNAP-IV total score, all the subgroups exhibited statistically significant improvement (younger group: from mean 1.30, SEM 0.05 to mean 1.09, SEM 0.06; T=4.44; n=70; *P*<.001; older group: from mean 1.39, SEM 0.11 to mean 1.11, SEM 0.09; T=2.91; n=27; *P*=.007; male group: from mean 1.35, SEM 0.05 to mean 1.12, SEM 0.05; T=4.78; n=86; *P*<.001; female group: from mean 1.14, SEM 0.08 to mean 0.87, SEM 0.16; T=2.59; n=11; *P*=.03 paired *t* test). No statistical difference was found between different age and gender groups (age: *P*=.56; gender: *P*=.74).

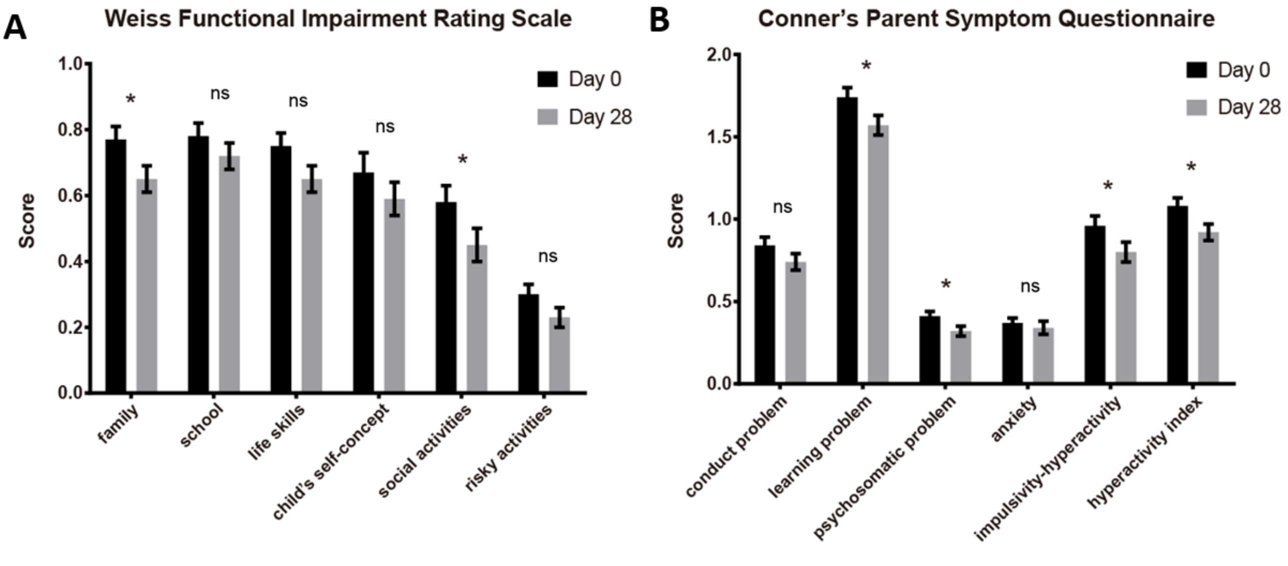
For the SNAP-IV AD score, only the younger group and male group exhibited statistically significant improvement (younger group: from mean 1.72, SEM 0.07 to mean 1.42, SEM 0.07; T=4.77; n=70; *P*<.001; male group: from mean 1.73, SEM 0.06 to mean 1.48, SEM 0.07; T=3.97; n=86; *P*<.001), while the older group and female group did not (older group: from mean 1.68, SEM 0.11 to mean 1.51, SEM 0.12; T=1.22; n=27; *P*=.23; female group: from mean 1.57, SEM 0.11 to mean 1.20, SEM 0.19; T=2.04; n=11; *P*=.07). No statistical difference was found between different age and gender groups (age: *P*=.39; gender: *P*=.57).

For the SNAP-IV HD score, the younger, older, and male groups exhibited statistically significant improvement (younger group: from mean 1.34, SEM 0.07 to mean 1.07, SEM 0.08; T=4.53; n=70; *P*<.001; older group: from mean 1.49, SEM 0.14 to mean 1.01, SEM 0.09; T=3.98; n=27; *P*<.001; male group: from mean 1.42, SEM 0.07 to mean 1.07, SEM 0.06; T=5.69; n=86; *P*<.001), while the female group did not (n=11; *P*=.09). No statistical difference was found between different age and gender groups (age: *P*=.94; gender: *P*=.26).

For the SNAP-IV ODD score, only the older group and female group exhibited statistically significant improvement (older group: from mean 0.95, SEM 0.13 to mean 0.76, SEM 0.11; T=2.50; n=27; *P*=.01; female group: from mean 0.70, SEM 0.10 to mean 0.47, SEM 0.08; T=2.02; n=11; *P*=.04), while the younger group and male group did not (younger group: T=1.38; n=70; *P*=.17; male group: T=1.88; n=86; *P*=.06). No statistical difference was found between different age and gender groups (age: *P*=.17; gender: *P*=.22).

The functional impairments of participants were measured using WFIRS (Figure 4A). WFIRS data did not pass the normality test; therefore, the Wilcoxon signed-rank test was used. We found that after 4 weeks of treatment, the impairments of the "family" and "social activities" domains were statistically significantly improved (family: from mean 0.75, SEM 0.05 to mean 0.65, SEM 0.04; *Z*=2.80; *P*=.01; social activities: from mean 0.56, SEM 0.05 to mean 0.45, SEM 0.05; *Z*=2.91; n=97; *P*=.01), while the domains of "school," "life skills," "child's self-concept," and "risky activities" exhibited nonsignificant improvement (school: *P*=.14; life skills: *P*=.09; child's self-concept: *P*=.34; risky activities: *P*=.14; n=97).

Figure 4. Statistical results of the Weiss Functional Impairment Rating Scale and Conner's Parent Symptom Questionnaire suggested significant improvement after 4 weeks of intervention. (A) Weiss Functional Impairment Rating Scale outcomes (scores) from day 0 to day 28. (B) Conner's Parent Symptom Questionnaire outcomes (scores) from day 0 to day 28. Error bars indicate SEM. ns: nonsignificant. *P<.05, **P<.01, ***P<.001.



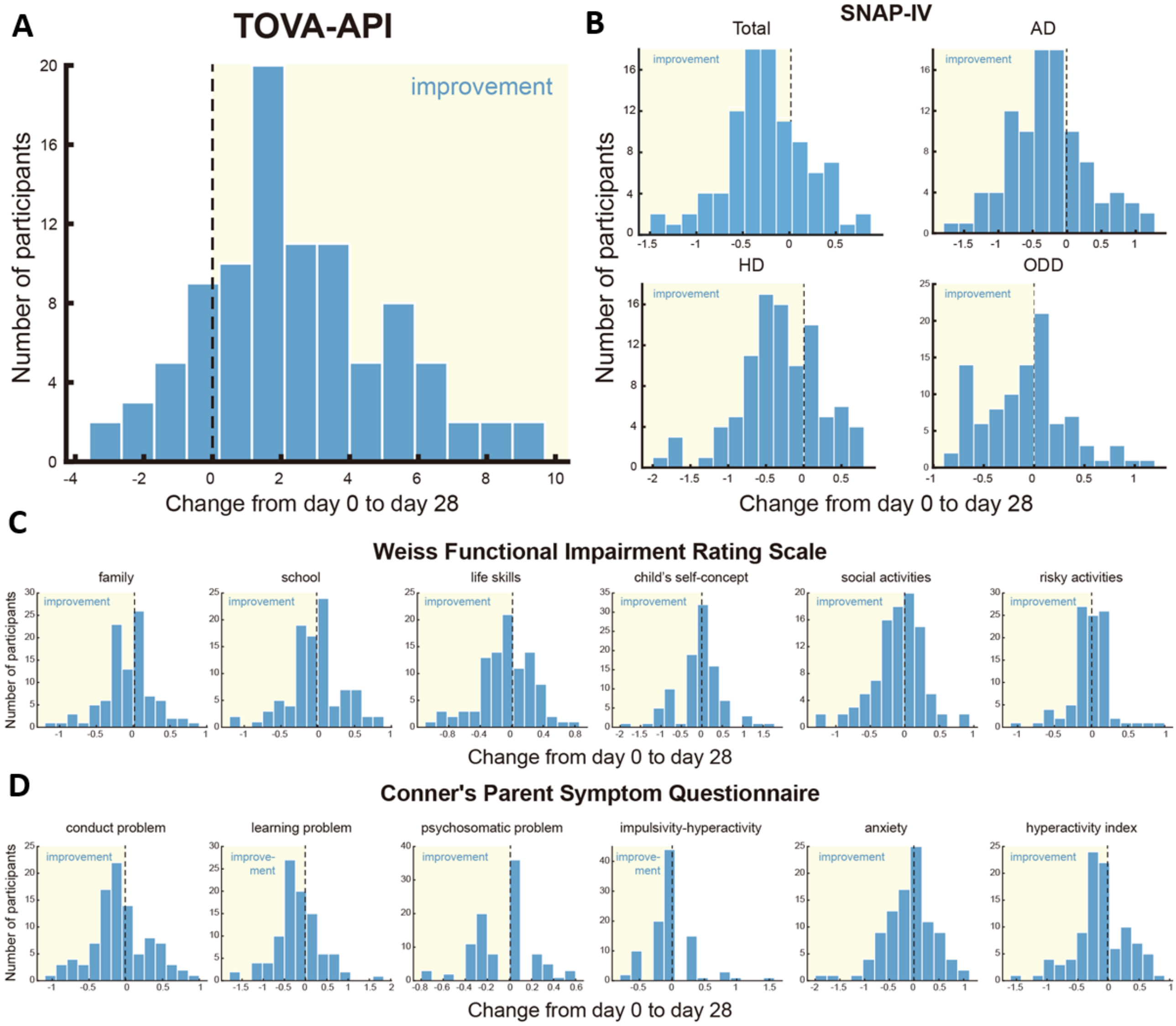
We also analyzed the number of items scored ≥2 in WFIRS to investigate the treatment effect on severe functional impairments using the chi-square test. We found that the correlation between the intervention and the number of items scored ≥2 of “child’s self-concept” was statistically significant (Pearson χ^2_1=7.249; *P*=.007). There were also almost significant correlations between the intervention and the number of items scored ≥2 of “family” and “life skills” (family: Pearson χ^2_1=3.183; *P*=.07; life skills: Pearson χ^2_1=3.071; *P*=.08). While there were no significant correlations between the intervention and the number of items scored ≥2 of “school,” “social activities,” and “risky activities” (school: Pearson χ^2_1=0.949; *P*=.33; social activities: Pearson χ^2_1=0.899; *P*=.34; risky activities: Pearson χ^2_1=0.519; *P*=.47).

The problematic behaviors of participants were measured using PSQ (Figure 4B). PSQ data did not pass the normality test; therefore, the Wilcoxon signed-rank test was used. After 4 weeks treatment, the domains of “learning problem,” “psychosomatic problem,” “impulsivity-hyperactivity,” and “hyperactivity index” were statistically significantly improved (learning problem: from mean 1.72, SEM 0.06 to mean 1.57, SEM 0.06; *Z*=2.42; *P*=.03; psychosomatic problem: from mean 0.40, SEM 0.03 to mean 0.32, SEM 0.03; *Z*=2.66; *P*=.02; impulsivity-hyperactivity: from mean 0.94, SEM 0.06 to mean 0.80, SEM 0.06; *Z*=2.49; *P*=.03; hyperactivity index: from mean 1.06, SEM 0.05 to mean 0.92, SEM 0.05; *Z*=2.90; *P*=.01; n=97). While the domains of “conduct problem” and “anxiety” exhibited nonsignificant improvement (conduct problem: *P*=.06; anxiety: *P*=.17).

The above findings suggested that the 4-week intervention could significantly improve the core symptoms of children with ADHD (measured by SNAP-IV), as well as a few domains related to functional impairments and problematic behaviors (measured by WFIRS and PSQ).

Table 2 summarizes the statistical results of efficacy outcomes during the 4-week treatment. The distributions of efficacy-outcome improvements (from day 0 to day 28) among the participants are shown in Figure 5. The proportions of participants whose efficacy outcomes showed improvements were 80% (76/95 in TOVA), 71.13% (69/97 in SNAP-IV total), 70.1% (68/97 in SNAP-IV AD), 70.1% (68/97 in SNAP-IV HD), 55.67% (54/97 in SNAP-IV ODD), 54.64% (53/97 in WFIRS-family), 52.58% (51/97 in WFIRS-school), 52.58% (51/97 in WFIRS-life skills), 38.14% (37/97 in WFIRS-child’s self-concept), 54.64% (53/97 in WFIRS-social activities), 42.27% (41/97 in WFIRS-risky activities), 58.76% (57/97 in PSQ-conduct problem), 48.45% (47/97 in PSQ-learning problem), 45.36% (44/97 in PSQ-psychosomatic problem), 32.99% (32/97 in PSQ-anxiety), 47.42% (46/97 in PSQ-impulsivity-hyperactivity), and 57.73% (56/97 in PSQ-hyperactivity index).

Figure 5. Distributions of efficacy-outcome improvements (from day 0 to day 28) among the participants suggest significant improvement of symptoms. (A) Distribution of Test of Variables of Attention (TOVA) Attention Performance Index (API) changes from day 0 to day 28; (B) Distributions of Swanson, Nolan, and Pelham Questionnaire, version 4 (SNAP-IV) score changes from day 0 to day 28; (C) Distributions of Weiss Functional Impairment Rating Scale score changes from day 0 to day 28; (D) Distributions of Conner's Parent Symptom Questionnaire score changes from day 0 to day 28. The ranges corresponding to improvement are highlighted. AD: inattention; HD: hyperactivity/impulsivity; ODD: oppositional defiant disorder.



Safety Results

Parents and participants were informed to report any safety concerns and discomfort during the study. Any concerns and discomfort reported by parents and participants during the study were recorded. Overall, no device-related AE or severe AE was observed or reported during the 4-week intervention and the 3-month follow-up period, suggesting the safety of treatment.

Discussion

Principal Findings

This paper reports the first Chinese clinical trial outcomes of an adaptive multitasking training paradigm for the treatment of children with ADHD without medication. This study preliminarily suggested the significant improvements of ADHD symptoms and attention function after 1-month digital therapy (combined with basic BPT, without medication) using the "Attention Enhancement Training Software" (ADHD-DTx), a video game–like training software for children with ADHD aged 6-12 years, with satisfying safety outcomes.

Comparison to Prior Work

The efficacy results are consistent with previous studies using an adaptive multitasking training paradigm [7,8,10], indicating the potential application value (as an individual or adjuvant treatment) of video game–like digital therapy for children with ADHD. The potential contribution of game-like cognitive training on attention has been reported both in children and adults [26-30], which suggests a new digital solution for attention-related deficits (such as ADHD, mild cognitive impairment, etc) across age groups [31,32].

The BPT has been extensively studied, and the overall effect size was estimated by previous researchers. Gubbels and colleagues [33] suggested that the overall effect of parent training was 0.416 (Cohen *d*) in their meta-analysis research, while Zwi and colleagues [23] suggested that the effect size might be "between small and medium" in their review. In comparison, we calculated the effect size (Cohen *d*) of the intervention in this study, based on TOVA-API and SNAP-IV data. We found that the effect size of TOVA-API was 0.80, while that of SNAP-IV total score was 0.50. These results exceeded the effect sizes of BPT estimated by previous researchers, suggesting that the ADHD-DTx digital therapeutic, when integrated with BPT, may yield superior outcomes to BPT alone. However, such cross-study comparisons are informal and cannot independently establish the efficacy of ADHD-DTx. More rigorously controlled trials—directly comparing the combined intervention against BPT alone—are needed to draw definitive conclusions regarding the stand-alone contribution of the digital therapy.

Compared with the FDA-approved EndeavorRx (AKL-T01) software, the "Attention Enhancement Training Software" (ADHD-DTx) added the "digit cancellation task" (a widely used attention assessment and training method in clinical practice [11]) to further enhance the training effect on attention function. The "digit cancellation task" was simplified from the clinical version and consisted of 5 different types of tasks: (1) choose a specific number (eg, "choose 3"), (2) choose the adjacent (left or right) number of a specific number (eg, "choose the adjacent number on the left of 3"), (3) choose a specific number adjacent (left or right) to a specific number (eg, "choose 2 on the left of 3"), (4) choose the number in the middle position of 2 specific numbers (eg, "choose the number in the middle position of 2 and 3"), and (5) choose a specific type (even or odd) of number in the middle position of 2 specific numbers (eg, "choose the even number in the middle position of 2 and 3"). These tasks targeted multiple dimensions of attention (such as pointing, shifting, selection, span, and allocation), aiming to activate extensive attention-related brain regions and achieve the training effect of attention function.

Limitations

The key objective of this study was to explore the clinical safety and efficacy of ADHD-DTx digital therapy. However, considering that this study was single-armed and did not use a control group, it would be meaningful to attempt to exclude the influence of basic BPT from the final therapeutic effect. We suggest that further research should use a randomized, double-blinded, parallel-controlled design, which could provide the most reliable data reflecting the actual effect of ADHD-DTx.

In this study, several potential confounding factors should be considered. First, the subjective evaluation bias was caused by the open-label design, the subjective parent-reported scales for all secondary outcomes, and the large research time span. Open-label design and subjective scales could introduce a significant risk of reporter bias or placebo effect, while the 1-year time span of this study might further enhance the potential subjective bias due to the fluctuating behavior patterns of children between school and vacation periods. These confounding factors could not be separated from the intervention's true effect, and future studies should adopt a double-blind, randomized controlled design to mitigate these biases.

Second, heterogeneity in the effectiveness of BPT may have contributed to variability in outcomes. The actual effect of BPT largely depended on the performance of parents when providing behavioral guidance to children with ADHD. However, due to differences in educational level, professional skills, and training quality, the actual effectiveness may vary among different parents, resulting in population heterogeneity. During the study, all parents were asked to daily report their home-based training progress, and all participants exhibited high adherence to both ADHD-DTx and BPT. And the subgroup analysis focusing on parental education level did not show a significant difference in efficacy outcome. Therefore, the potential influence of adherence and parental education level may be limited. However, the potential impact of training quality differences between parents still existed.

Third, the potential contribution of BPT was not excluded in this single-arm study design, which might amplify the effectiveness results. As a standard foundational treatment for ADHD, BPT was mandated by the regulatory agency for all participants due to ethical considerations. While its efficacy was well-established, albeit generally mild, its concurrent implementation with the investigational digital therapy (ADHD-DTx) introduced a considerable confounding effect that may inflate the perceived effectiveness of the intervention. Improved study design should include a control group (providing BPT equally to both test and control groups, while only the test group is treated with ADHD-DTx) so as to effectively control the confounding influence of BPT, allowing for a more precise assessment of the benefit of ADHD-DTx.

Fourth, the generalizability of the results was constrained due to limited sample diversity. The included participants were predominantly male (99/110, 90%) and of Han ethnicity (106/110, 96.36%), which weakened the interpretability of outcomes for females and other ethnicities. Future research should prioritize enrolling more representative samples with balanced gender distribution and greater ethnic diversity to ascertain the efficacy of the intervention across the broader population with ADHD.

Fifth, the rigorously monitored and controlled conditions of a clinical trial may not perfectly predict the intervention's effectiveness in real-world clinical practice. The high adherence observed in this study was facilitated by intensive monitoring and support, which limited the direct translation of these efficacy results into real-world practice, where effectiveness might be weaker. Consequently, future investigation should involve pragmatic trials to generate robust evidence on the actual efficacy and implementation of the intervention.

Finally, the subgroup analyses for age and gender were likely underpowered due to the small sample sizes in the female and older groups. Consequently, any observed differences (or lack thereof) between these subgroups should be considered preliminary and interpreted with caution. Further research specifically designed to investigate these demographic groups is needed to advance our understanding of digital therapeutics for ADHD.

Future Directions

Further improvements could still be carried out in future research. In this single-arm, open-label study, we preliminarily revealed the safety and efficacy of ADHD-DTx digital therapy. While future research should use a randomized, double-blinded, parallel controlled design so as to prevent potential placebo effect and rule out the effect of basic BPT at home, it will achieve a better evaluation of the actual effect size of ADHD-DTx. Further study could also investigate the effect of a prolonged treatment period (>4 weeks) to study the long-term influence of digital therapy. Longer safety monitoring after the intervention period is also needed to reveal potential effects on visual acuity, daily activities, sleep quality, screen usage time, and so on. The independent ethical oversight should continue to be valued and implemented to avoid potential conflicts of interest. Future studies should include comparisons with well-established interventions like pharmacotherapy, psychological behavioral therapy, and so on, to obtain more informative evidence for better clinical practice.

There is psychological and neurophysiological evidence suggesting the potential mechanisms of ADHD treatment by the adaptive multitasking paradigm. According to electroencephalogram studies focusing on the neural mechanisms of the adaptive multitasking paradigm, after 4 weeks of training, participants' FM significantly enhanced [9,10]. The FM is a well-established neural marker of attention control [34-38], and its enhancement is related to the suppression of a key node of the "default mode network" [10,39,40], leading to a reduction of the susceptibility to internal distraction, resulting in better task performance and sustained attention. To explore potential neural mechanisms of digital therapy, physiological measurements of brain function (such as electroencephalogram, functional magnetic resonance imaging, functional near-infrared spectroscopy, functional ultrasound, etc) could provide irreplaceable insight and should be considered in future studies.

The evaluation methods used in this study were limited, and further study could include other powerful tools such as other neuropsychological tests (eg, the Integrated Visual and Auditory Continuous Performance Test, the Wisconsin Card Sorting Test, the Cambridge Neuropsychological Test Automatic Battery, etc), classic scales (eg, the Execution Function Parent Questionnaire, the Vanderbilt ADHD Parent Rating Scale, etc).

Due to the limited data on female patients, we suggest that further research could include a higher proportion of female patients so as to make the results more representative. The inclusion and exclusion criteria set strict restrictions on comorbidity, which, on the one hand, could reduce potential confounding factors; however, on the other hand, it would limit our knowledge of real-world situations. We suggest that subsequent research can specifically investigate the effect of ADHD digital therapy involving comorbidities. We also suggest including the examination of visual acuity in further study to investigate the potential impact of long-term electronic device training on vision.

Conclusions

According to our pilot study on the Attention Enhancement Training Software (ADHD-DTx), we suggest that it could be used as a daily home-training tool for children with ADHD as an adjunct therapy to medication or other behavioral therapies, thereby further improving the intervention efficacy and reducing potential adverse drug reactions.

Acknowledgments

No artificial intelligence–generated text was used in this paper.

Funding

We thank all volunteers for their participation in the study. This work was supported in part by STI 2030-Major Projects (grant 2021ZD0200400), in part by the National Natural Science Foundation of China (grant 62336007), in part by the Key Research and Development Program of Zhejiang (grant 2023C03003), in part by the Key R&D Program of Zhejiang (2024SSYS0016), in part by the Starry Night Science Fund of the Zhejiang University Shanghai Institute for Advanced Study (grant SN-ZJU-SIAS-002), in part by the Fundamental Research Funds for the Central Universities, in part by ZJU-GENSCI Children's Health Research & Development Center (ZJU-GENSCI2024YB003), in part by the Project for Hangzhou Medical Disciplines of Excellence, and in part by the Key Project for Hangzhou Medical Disciplines.

Data Availability

The data that support the findings of this study are available on request from the corresponding author upon reasonable request.

Authors' Contributions

LY, KJ, and WL designed the study. MB, JY, Jiangping Wang, and KJ collected the data. Jiaheng Wang and SF contributed unpublished analytic scripts. SF analyzed the data and drafted the manuscript.

Conflicts of Interest

SF and WL are employed by SDO Digital Therapeutics (a provider of advanced digital and neurological therapies for children) at the time of submission, while SF did not have any financial or nonfinancial competing interests during the clinical trial period

(2021). SDO Digital Therapeutics did not participate in the data collection of this study. All other authors declare no financial or nonfinancial competing interests. The views and opinions expressed within this manuscript are those of all the authors. Our adherence to the policies on data availability is not altered.

Multimedia Appendix 1
Complete inclusion and exclusion criteria.
[DOCX File , 17 KB - games_v14i1e76114_app1.docx]

Multimedia Appendix 2
Description of outcome measurements.
[DOCX File , 14 KB - games_v14i1e76114_app2.docx]

References

1. Faraone SV, Banaschewski T, Coghill D, Zheng Y, Biederman J, Bellgrove MA, et al. The World federation of ADHD international consensus statement: 208 evidence-based conclusions about the disorder. Neurosci Biobehav Rev 2021;128:789-818 [FREE Full text] [doi: 10.1016/j.neubiorev.2021.01.022] [Medline: 33549739]
2. Dalsgaard S, Mortensen PB, Frydenberg M, Thomsen PH. ADHD, stimulant treatment in childhood and subsequent substance abuse in adulthood - a naturalistic long-term follow-up study. Addict Behav 2014;39(1):325-328. [doi: 10.1016/j.addbeh.2013.09.002] [Medline: 24090624]
3. Pelham WE, Fabiano GA, Waxmonsky JG, Greiner AR, Gnagy EM, Pelham WE, et al. Treatment sequencing for childhood ADHD: a multiple-randomization study of adaptive medication and behavioral interventions. J Clin Child Adolesc Psychol 2016;45(4):396-415 [FREE Full text] [doi: 10.1080/15374416.2015.1105138] [Medline: 26882332]
4. Qiu H, Liang X, Wang P, Zhang H, Shum DH. Efficacy of non-pharmacological interventions on executive functions in children and adolescents with ADHD: a systematic review and meta-analysis. Asian J Psychiatr 2023;87:103692 [FREE Full text] [doi: 10.1016/j.ajp.2023.103692] [Medline: 37450981]
5. He F, Qi Y, Zhou Y, Cao A, Yue X, Fang S, et al. Meta-analysis of the efficacy of digital therapies in children with attention-deficit hyperactivity disorder. Front Psychiatry 2023;14:1054831 [FREE Full text] [doi: 10.3389/fpsyt.2023.1054831] [Medline: 37260755]
6. Davis NO, Bower J, Kollins SH. Proof-of-concept study of an at-home, engaging, digital intervention for pediatric ADHD. PLoS One 2018;13(1):e0189749 [FREE Full text] [doi: 10.1371/journal.pone.0189749] [Medline: 29324745]
7. Kollins SH, DeLoss DJ, Cañadas E, Lutz J, Findling RL, Keefe R, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. Lancet Digit Health 2020;2(4):e168-e178 [FREE Full text] [doi: 10.1016/S2589-7500(20)30017-0] [Medline: 33334505]
8. Kollins SH, Childress A, Heusser AC, Lutz J. Effectiveness of a digital therapeutic as adjunct to treatment with medication in pediatric ADHD. NPJ Digit Med 2021;4(1):58 [FREE Full text] [doi: 10.1038/s41746-021-00429-0] [Medline: 33772095]
9. Gallen CL, Anguera JA, Gerdes MR, Simon AJ, Cañadas E, Marco EJ. Enhancing neural markers of attention in children with ADHD using a digital therapeutic. PLoS ONE 2021;16(12):e0261981. [doi: 10.1371/journal.pone.0261981] [Medline: 34972140]
10. Anguera JA, Boccanfuso J, Rintoul JL, Al-Hashimi O, Faraji F, Janowich J, et al. Video game training enhances cognitive control in older adults. Nature 2013;501(7465):97-101. [doi: 10.1038/nature12486] [Medline: 24005416]
11. Hatta T, Yoshizaki K, Ito Y, Mase M, Kabasawa H. Reliability and validity of the digit cancellation test, a brief screen of attention. Int J Psychol Sci 2012;55(4):246-256. [doi: 10.2117/psysoc.2012.246]
12. van den Hoofdakker BJ, van der Veen-Mulders L, Sytema S, Emmelkamp PMG, Minderaa B, Nauta MH. Effectiveness of behavioral parent training for children with ADHD in routine clinical practice: a randomized controlled study. J Am Acad Child Adolesc Psychiatry 2007;46(10):1263-1271. [doi: 10.1097/chi.0b013e3181354bc2] [Medline: 17885567]
13. Chacko A, Jensen SA, Lowry LS, Cornwell M, Chimklis A, Chan E, et al. Engagement in behavioral parent training: review of the literature and implications for practice. Clin Child Fam Psychol Rev 2016;19(3):204-215. [doi: 10.1007/s10567-016-0205-2] [Medline: 27311693]
14. Forbes GB. Clinical utility of the test of variables of attention (TOVA) in the diagnosis of attention-deficit/hyperactivity disorder. J Clin Psychol 1998;54(4):461-476. [doi: 10.1002/(sici)1097-4679(199806)54:4<461::aid-jclp8>3.0.co;2-q]
15. Greenberg LM, Waldman ID. Developmental normative data on the test of variables of attention (T.O.V.A.). J Child Psychol Psychiatry 1993;34(6):1019-1030. [doi: 10.1111/j.1469-7610.1993.tb01105.x] [Medline: 8408366]
16. Swanson J. School-Based Assessments and Interventions for ADD Students. Frederick, Maryland, USA: KC publishing; 1992.
17. Swanson JM, Kraemer HC, Hinshaw SP, Arnold LE, Conners CK, Abikoff HB, et al. Clinical relevance of the primary findings of the MTA: success rates based on severity of ADHD and ODD symptoms at the end of treatment. J Am Acad Child Adolesc Psychiatry 2001;40(2):168-179. [doi: 10.1097/00004583-200102000-00011] [Medline: 11211365]

18. Swanson JM, Sandman CA, Deutsch C, Baren M. Methylphenidate hydrochloride given with or before breakfast: I. behavioral, cognitive, and electrophysiologic effects. Pediatrics 1983;72(1):49-55. [Medline: 6866591]
19. Weiss MD, McBride NM, Craig S, Jensen P. Conceptual review of measuring functional impairment: findings from the Weiss Functional Impairment Rating Scale. Evid Based Ment Health 2018;21(4):155-164 [FREE Full text] [doi: 10.1136/ebmental-2018-300025] [Medline: 30314990]
20. Fan J, Du Y, Wang L. The norm and reliability of the Conners Parent Symptom Questionnaire in Chinese urban children. Shanghai Arch Psychiatry 2005;17(6):321-323.
21. Conners CK, Sitarenios G, Parker JDA, Epstein JN. The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. J Abnorm Child Psychol 1998;26(4):257-268. [doi: 10.1023/a:1022602400621] [Medline: 9700518]
22. van der Oord S, Tripp G. How to improve behavioral parent and teacher training for children with ADHD: integrating empirical research on learning and motivation into treatment. Clin Child Fam Psychol Rev 2020;23(4):577-604 [FREE Full text] [doi: 10.1007/s10567-020-00327-z] [Medline: 32968886]
23. Zwi M, Jones H, Thorgaard C, York A, Dennis JA. Parent training interventions for attention deficit hyperactivity disorder (ADHD) in children aged 5 to 18 years. Cochrane Database Syst Rev 2011;2011(12):CD003018 [FREE Full text] [doi: 10.1002/14651858.CD003018.pub3] [Medline: 22161373]
24. Hornstra R, Groenman AP, van der Oord S, Luman M, Dekkers TJ, van der Veen-Mulders L, et al. Review: which components of behavioral parent and teacher training work for children with ADHD? - a metaregression analysis on child behavioral outcomes. Child Adolesc Ment Health 2023;28(2):258-268. [doi: 10.1111/camh.12561] [Medline: 35417075]
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 1995;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
26. Anguera JA, Schachtner JN, Simon AJ, Volponi J, Javed S, Gallen CL, et al. Long-term maintenance of multitasking abilities following video game training in older adults. Neurobiol Aging 2021;103:22-30. [doi: 10.1016/j.neurobiolaging.2021.02.023] [Medline: 33789209]
27. Jurigova BG, Gerdes MR, Anguera JA, Marco EJ. Sustained benefits of cognitive training in children with inattention, three-year follow-up. PLoS One 2021;16(2):e0246449 [FREE Full text] [doi: 10.1371/journal.pone.0246449] [Medline: 33539468]
28. Anguera JA, Volponi JJ, Simon AJ, Gallen CL, Rolle CE, Anguera-Singla R, et al. Integrated cognitive and physical fitness training enhances attention abilities in older adults. NPJ Aging 2022;8(1):12 [FREE Full text] [doi: 10.1038/s41514-022-00093-y] [Medline: 36042247]
29. Gallen CL, Schachtner JN, Anguera-Singla R, Anguera JA, Gazzaley A. Influence of game features on attention in adults. Front Psychol 2023;14:1123306 [FREE Full text] [doi: 10.3389/fpsyg.2023.1123306] [Medline: 37228349]
30. Ziegler DA, Simon AJ, Gallen CL, Skinner S, Janowich JR, Volponi JJ, et al. Closed-loop digital meditation improves sustained attention in young adults. Nat Hum Behav 2019;3(7):746-757 [FREE Full text] [doi: 10.1038/s41562-019-0611-9] [Medline: 31160812]
31. Mishra J, Anguera J, Gazzaley A. Video games for neuro-cognitive optimization. Neuron 2016;90(2):214-218 [FREE Full text] [doi: 10.1016/j.neuron.2016.04.010] [Medline: 27100194]
32. Anguera JA, Gazzaley A. Video games, cognitive exercises, and the enhancement of cognitive abilities. Curr Opin Behav Sci 2015;4:160-165. [doi: 10.1016/j.cobeha.2015.06.002]
33. Gubbels J, van der Put CE, Assink M. The effectiveness of parent training programs for child maltreatment and their components: a meta-analysis. Int J Environ Res Public Health 2019;16(13):2404 [FREE Full text] [doi: 10.3390/ijerph16132404] [Medline: 31284575]
34. Cavanagh JF, Frank MJ. Frontal theta as a mechanism for cognitive control. Trends Cogn Sci 2014;18(8):414-421 [FREE Full text] [doi: 10.1016/j.tics.2014.04.012] [Medline: 24835663]
35. Clayton MS, Yeung N, Cohen Kadosh R. The roles of cortical oscillations in sustained attention. Trends Cogn Sci 2015;19(4):188-195. [doi: 10.1016/j.tics.2015.02.004] [Medline: 25765608]
36. Brandmeyer T, Delorme A. Closed-loop frontal midlineθ neurofeedback: a novel approach for training focused-attention meditation. Front Hum Neurosci 2020;14:246 [FREE Full text] [doi: 10.3389/fnhum.2020.00246] [Medline: 32714171]
37. Messel MS, Raud L, Hoff PK, Stubberud J, Huster RJ. Frontal-midline theta reflects different mechanisms associated with proactive and reactive control of inhibition. Neuroimage 2021;241:118400 [FREE Full text] [doi: 10.1016/j.neuroimage.2021.118400] [Medline: 34311382]
38. Cowley BU, Juurmaa K, Palomäki J. Reduced power in fronto-parietal theta EEG linked to impaired attention-sampling in adult ADHD. eNeuro 2022;9(1) [FREE Full text] [doi: 10.1523/ENEURO.0028-21.2021] [Medline: 34893505]
39. Tu W, Ma Z, Ma Y, Dopfel D, Zhang N. Suppressing anterior cingulate cortex modulates default mode network and behavior in awake rats. Cereb Cortex 2021;31(1):312-323 [FREE Full text] [doi: 10.1093/cercor/bhaa227] [Medline: 32820327]
40. Chadick JZ, Zanto TP, Gazzaley A. Structural and functional differences in medial prefrontal cortex underlie distractibility and suppression deficits in ageing. Nat Commun 2014;5:4223 [FREE Full text] [doi: 10.1038/ncomms5223] [Medline: 24979364]

Abbreviations

AD: inattention subscale of SNAP-IV
ADHD: attention-deficit/hyperactivity disorder
ADHD-RS-IV: Attention-Deficit/Hyperactivity Disorder Rating Scale IV
AE: adverse event
API: Attention Performance Index
BPT: behavioral parent training
DSM-V: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
FDA: US Food and Drug Administration
FM : frontal midline theta
HD: hyperactivity/impulsivity subscale of SNAP-IV
NMPA: National Medical Products Administration
ODD: oppositional defiant disorder subscale of SNAP-IV
PSQ: Conner's Parent Symptom Questionnaire
SEM: SE of the mean
SNAP-IV: Swanson, Nolan, and Pelham Questionnaire, version 4
TOVA: Test of Variables of Attention
WFIRS: Weiss Functional Impairment Rating Scale

Edited by A Coristine; submitted 16.Apr.2025; peer-reviewed by S Mehan, S Triscari; comments to author 27.Jun.2025; accepted 22.Oct.2025; published 05.Jan.2026.

Please cite as:
Bao M, Feng S, Wang J, Ye J, Wang J, Li W, Jiang K, Yao L
Efficacy and Safety of a Video Game–Like Digital Therapy Intervention for Chinese Children With Attention-Deficit/Hyperactivity Disorder: Single-Arm, Open-Label Pre-Post Study
JMIR Serious Games 2026;14:e76114
URL: https://games.jmir.org/2026/1/e76114
doi:10.2196/76114
PMID:

Original Paper

Gamified Physical Education and Cognitive Performance Among Chinese Secondary School Students: Cross-Sectional Moderation Mediation Study

Jingtong Zhang[1], PhD; Sae-Sook Oh[1], Prof Dr; Yaru Xu[2], MSc

[1]Department of Physical Education, Kyonggi University, Suwon-si, Republic of Korea

[2]School of Health Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia

Corresponding Author:
Yaru Xu, MSc
School of Health Sciences
Universiti Sains Malaysia
Exercise and Sports Science, School of Health Sciences
Universiti Sains Malaysia, Health Campus
Kubang Kerian, Kelantan, 16150
Malaysia
Phone: 86 17763575813
Email: xuyarustc@163.com

Abstract

Background: Educators are exploring new methods to educate beyond the classroom as global concerns about students' cognitive, emotional, and social well-being grow. Physical education (PE) has been demonstrated to boost cognitive and psychological outcomes in several studies. Most research has neglected the benefits of gamification and artificial intelligence (AI)–based feedback in PE, focusing instead on conventional PE formats. The impacts of technologically enhanced PE settings on students' cognitive performance through feedback and reward mechanisms remain understudied.

Objective: This study aimed to investigate how intrinsic motivation and AI-based feedback moderated the effects of gamified PE on students' cognitive performance.

Methods: The study used a cross-sectional design. In Beijing, Shanghai, Chengdu, and Guangzhou, a total of 1029 public high school students completed a standardized questionnaire. Students in secondary school (male: n=490, 47.6% and female: n=539, 52.4%) aged 10-18 years, were recruited from urban, suburban, and rural locales. Participants were sourced from public, private, and semigovernment schools, reflecting a range of academic achievement levels and access to technology. Students participating in standard PE sessions were included, whereas those with medical conditions that restricted physical exercise were excluded. Data were gathered via standardized questionnaires during designated PE sessions. Gamified PE, cognitive performance, intrinsic motivation, teacher support, collaboration, and AI feedback were examined using standardized instruments. Trained facilitators helped younger participants understand and follow ethical norms. The study used maximum likelihood estimation for structural equation modeling. Bootstrapping was used to analyze mediation and moderation effects at a 5% significance level (α=.05).

Results: According to structural equation modeling, gamified PE highly predicts cognitive performance (β=.34; *P*<.001). Other significant factors were teacher support (β=.31; *P*<.001), physical exercise enjoyment (β=.28; *P*<.001), and teamwork (β=.26; *P*<.001). AI-based feedback strengthened the link between gamified PE and cognitive performance under moderation analysis (β=.18; *P*<.001). Mediation analysis indicated that intrinsic motivation partially mediated the relationship, resulting in a significant indirect effect (β=.21, 95% CI 0.12-0.31; SE=0.05).

Conclusions: This research integrates gamified PE with AI-based feedback mechanisms to evaluate students' cognitive outcomes, a domain that has been rarely investigated experimentally. This study highlights the combined effect of intrinsic motivation and AI-generated feedback in a technology-enhanced PE context, in contrast to previous research that primarily focuses on traditional PE methods or isolated gamification elements. The findings enhance the field by demonstrating that student-centered, feedback-rich PE environments may improve cognitive abilities through social interaction, enjoyment, and instructor support. AI-assisted, gamified PE programs may enhance learning outcomes and academic performance among secondary school students.

(JMIR Serious Games 2026;14:e81086) doi:10.2196/81086

KEYWORDS

gamified physical education; cognitive performance; intrinsic motivation; AI-based feedback; teacher support; peer collaboration; China; artificial intelligence

Introduction

Background of Research

In the 21st century, multimodal techniques that integrate physical, cognitive, emotional, and technical components have become more important in education [1]. Contemporary research suggests that physical and intellectual education must be integrated, even though traditional schools generally divide them [2,3]. Physical education (PE) can change children's lives by enhancing their emotional, cognitive, and academic performance [4]. Concerns about student inactivity and cognitive overload have led to the development of gamified PE, an innovative teaching approach. This strategy uses game-design principles to make exercise more enjoyable, goal-oriented, and cognitively demanding [5]. Gaming methods combined with physical exercise have been shown to enhance motivation, teamwork, and self-control—essential for both mental and physical well-being [6].

Gamification incorporates features such as points, competition, prizes, badges, levels, and leaderboards into nongame contexts to motivate and engage users. Gamification, which originated in marketing and business, is now being used in classrooms to engage students [7]. This type of PE transforms static exercises into game-like cooperative learning settings where students work together to achieve goals. Research suggests that gamified methods improve students' engagement, enjoyment, and academic success [8]. Teenagers often find PE uninteresting or irrelevant; therefore, these strategies have been effective in capturing their attention. Gamified PE promotes active participation and cognitive engagement by allowing students to make choices, solve problems, and reflect on outcomes in real time [9].

Cognitive performance, encompassing attention, memory, reasoning, and problem-solving, has been extensively studied in educational psychology about physical well-being [10,11]. Regular exercise improves working memory, executive function, and information processing speed. Neurobiological mechanisms, including increased cerebral blood flow, reduced cortisol levels, and the release of brain-derived neurotrophic factor, may explain this connection [12]. Gamified PE may directly stimulate cognitive brain networks through rule-following, decision-making, instruction, memory, and strategy adjustment [13]. China's rapidly changing educational system is a fascinating environment for studying these dynamics. China has put student well-being alongside academic success in recent years to address excessive academic pressure, mental health issues, and physical inactivity in school-age children and adolescents [14]. National educational reforms aim to provide students with a "quality education" that nurtures their emotional, physical, and cognitive potential [15]. Chinese schools are experimenting with gamified teaching, notably in PE, to achieve this goal [16]. Despite well-defined policy aims, empirical research on the implementation processes and outcomes of these innovations is lacking. Understanding the impact of gamified PE on cognitive development and academic achievement in Chinese children requires considering psychological factors, such as motivation, and external factors, such as technology use.

Intrinsic motivation links gamified PE to cognitive performance. This is the desire to do something because it is fun, fascinating, or valuable. Ryan and Deci's [17] self-determination theory states that intrinsic motivation is driven by autonomy, competence, and relatedness. Gamified situations provide autonomy, competence, and relatedness. Gamified aspects in PE may help students understand and connect with physical exercise by making repetitive tasks more interesting [18]. Intrinsic motivation enables students to self-regulate their participation, thereby improving cognitive functioning and task performance [19]. The mediating role of intrinsic motivation is crucial in understanding how gamified PE may impact cognitive outcomes. Gamified PE may be enhanced or mitigated by technology, particularly artificial intelligence (AI)–powered feedback systems. Many progressive schools worldwide use smart wearables, mobility trackers, and digital dashboards, with AI being increasingly integrated [20,21]. These technologies enhance training efficiency by offering adaptive learning environments, personalized feedback, and real-time performance monitoring. AI may transform PE by enabling real-time monitoring of students' and teachers' efficiency, speed, endurance, and progress [22]. By acknowledging effort, fostering reflection, and driving self-improvement, such technologies may increase student engagement when gamified. Student receptivity to technology, teacher implementation skills, and institutional resources can all impact their effectiveness. Thus, AI-based feedback may moderate the relationship between gamified PE and cognitive performance sufficiently to change its direction.

In addition to these core concepts, contextual considerations should also be addressed. In China, a collectivist culture that values group cohesiveness and cooperative learning, peer collaboration is important. Gamified PE involves collaborative, communicative, and goal-oriented teamwork [23]. This may improve relationships and higher-order thinking via collaborative problem-solving and strategy building. Teacher support also affects student engagement in gamified contexts. When teachers model healthy exercise habits, provide constructive feedback, and encourage autonomy, they may motivate and engage students. Instructor behavior and gamified content delivery are crucial factors in assessing the effectiveness of an intervention [24]. Another key component in China's educational background is the variety of school settings, from public to private and urban to rural. The accessibility of gamification and AI technologies may vary. In Shanghai and Beijing, schools have smart classrooms and instructors who use technology to make learning an engaging experience [25]. Rural or low-resource schools often struggle with outdated technology, a shortage of qualified personnel, and resistance to innovative teaching methods. Demographics and institutional characteristics may help explain differences in student cognitive performance. Confucian values, including diligence, self-control, and academic achievement, also shape Chinese educational perspectives [26]. These principles influence students' perceptions of gamification and other innovative teaching methods. Gamified PE may be invigorating for some students but too limited for others. Thus, cultural perception is a subtle yet powerful factor that may influence new teaching approaches [27]. To achieve cognitive and motivational advantages, gamified PE may need culturally relevant values and communication strategies.

This study should be viewed from the perspective of modern educational research, which increasingly promotes interdisciplinary collaboration. Motivation theory, PE, educational technology, and cognitive science must collaborate to understand student learning. Gamified PE combines movement studies, behavioral psychology, and technology. These cross-domain interactions are seldom studied, particularly in China's sociocultural context. Gamification in academic fields such as mathematics or language learning has garnered the greatest attention [28], whereas most research has focused on the cognitive advantages of physical exercise [29]. Technical, motivational, and cognitive studies on gamified PE as a core intervention are lacking.

This study is theoretically and practically important for Chinese educational innovation and student growth. Gamifying PE programs can improve children's health and engagement, a trend that is becoming increasingly significant in schools worldwide [30]. Gamified PE may increase physical activity and cognitive function in China, where academic achievement often takes precedence over emotional and physical growth [31]. Studying the mediating role of intrinsic motivation reveals the psychological underpinnings behind student learning and engagement. While educational technology shapes learning settings, AI-based feedback serves as a moderating component that addresses this tendency. This study's results contribute to the global discussion on educational reform, technology-enhanced learning, and student motivation, as it is one of the first to systematically analyze these factors within the Chinese secondary school system. New technologies help school administrators, policymakers, and educators develop more engaging, productive, and intellectually stimulating PE programs.

Today's educational studies must grasp how PE, cognitive development, and motivational psychology interact, particularly in rapidly modernizing nations such as China. Given the growing use of gamified methods and AI in education, research on their influence on key learning outcomes is urgently needed. Over the past 20 years, research on the cognitive benefits of exercise has evolved. Still, the impact of 2 modern pedagogical innovations, namely gamification and AI-based feedback, on this area remains unknown. This study emphasizes gamified PE, intrinsic motivation, AI-based feedback, and cognitive performance. The 3subsections of the review conclude with a hypothesis to guide empirical research.

Gamified PE and Physical Activity Enjoyment With Cognitive Performance

Several studies have demonstrated that gamified PE enhances children's learning and brain development [30,32]. Rule-based, competitive, and strategic games boost students' attention, engagement, and decision-making in gamified PE [33]. Gamified learning activities have been shown to improve task completion and working memory. Gamified physical activities compel students to solve problems, devise strategies, and assess their progress in real time, which promotes higher-order thinking. These benefits stem from PE games and competitions that keep children moving and engaged in critical-thinking activities [34]. Gamified PE lessons improved content retention and self-regulated learning. It is widely recognized that exercise enhances cognitive function. Students who like exercising are more likely to persist with it. This aids brain growth and executive function [35,36]. A meta-analysis found that children who were happier while exercising performed better on cognitive flexibility and memory tests. Engaging and pleasant activities promote academic learning, critical dual-task performance, and prefrontal brain activity. Despite a lack of scientific data, gamification is gaining popularity in China as a teaching method [37,38]. Gamified PE treatments in Shanghai, China, have improved students' memory and attention. Gamified PE increased mental rotation exercises and digit span recall. These results support the premise that engaging and interesting physical exercise promotes both physical and psychological health. Gamification's novelty must be managed carefully to keep users interested and minimize cognitive fatigue [39,40].

Peer Collaboration and Teacher Support to Cognitive Performance

Working together in class has long been shown to boost students' social and cognitive development. Collaborative learning plays a key role in intellectual growth. Teamwork and collaboration in PE foster effective communication, critical analysis, and perspective-taking [41,42]. Students who worked together on physical activities performed better on reflective thinking and adaptation exams. Peer contact in PE improves strategic thinking and decision-making, particularly when participants must collaborate to solve challenges or reach a consensus [43,44]. Coordinated peer collaboration in PE enhances cognitive engagement. This was especially evident when students discussed approaches and reflected on their outcomes. Cognitive benefits have been observed in teaching games for understanding models, which rely heavily on peer interactions [45,46]. These models improve foresight, planning, and problem-solving. Collectivist principles that promote group cohesiveness and shared accomplishment make peer collaboration meaningful in Chinese education. Chinese students who took PE programs with a partner performed better on executive functioning assessments. Peer collaboration and teacher support help influence students' cognitive involvement [47]. Teachers who foster autonomy and provide clear feedback help pupils build intrinsic motivation and learning perseverance. PE teachers' primary duties include engaging students and making the classroom exciting. Scaffolding, encouragement, and feedback improve cognitive performance [48,49]. Physical ability and academic success were higher in Chinese students who were more encouraged by their PE teachers. Formative assessments and motivational signals reduced cognitive load and increased self-reflection in PE students [50]. Gamified PE highlighted the need for psychologically safe spaces where children can experiment, fail, and learn, which is essential for cognitive development [51]. Instructors remain crucial to PE courses that use AI to support students in understanding digital feedback and to foster metacognition [52]. These findings suggest that students need to collaborate and have their teachers' support to increase cognitive performance, particularly in dynamic and game-based learning environments.

Mediating Role of Intrinsic Motivation and Moderating Effect of AI-Based Feedback

Intrinsic motivation, which is the drive to do something for its own sake, has been linked to scholastic and cognitive success. Gamified learning settings make students feel more connected, competent, and independent [53]. Motivation to study enhances students' use of deeper cognitive techniques, sustained attention, and effort. Genuinely driven PE students performed better and persevered longer in motor learning tasks. Enjoyment and perceived competence during PE influenced the relationship between PE involvement and academic success [54,55]. Inner motivation mediated the relationship between physical activity and math performance in Chinese middle schools. A recent study suggests that gamified PE motivates students to enhance their executive functioning and facilitate learning transfer [56,57]. These findings support the cognitive benefits of gamified PE, including increased intrinsic motivation, perceived competence, and a sense of control. In contrast, AI has enhanced PE by providing students with real-time performance statistics. Smart wristbands equipped with AI-powered motion sensors facilitate reflective thinking and personalized learning. AI-based feedback enhanced students' metacognitive awareness and adaptive learning strategies [58,59]. This input may reinforce learning loops, give immediate incentives, and drive persistent participation in gamified contexts. Students who received AI-enhanced PE courses performed better on spatial thinking and problem-solving examinations than those who received conventional instructor feedback. AI's moderating influence is not always positive. Students may become excessively dependent on external validation or overwhelmed by AI technology, depending on deployment and preparation [60,61]. Thus, its effectiveness may vary depending on the situation. AI-enhanced PE pilot programs in China have shown promising outcomes, with improved engagement and physical literacy [62]; however, concerns remain regarding unequal access and teacher training. When paired with AI-based feedback, gamified PE may improve cognitive function, particularly depending on the quality of feedback, engagement, and customization.

Research Gaps and Contribution of the Study

There is growing evidence that PE is important for cognitive development; however, the impact of gamified PE interventions on cognitive outcomes remains unclear, particularly in the Chinese educational and cultural context. Exercise has been shown to improve attention, memory, and executive functioning in several studies [63,64]. These studies have primarily examined aerobic or endurance activities, rather than PE programs that use games to enhance these areas. Gkintoni et al [65] found that motor skill memorization in conventional PE improves mental agility, while gamified activities that require flexibility, strategy, and decision-making in the present engage students' brains more effectively.

Another topic with limited research is the impact of PE on cognitive functioning and its underlying mechanisms. Self-determination theory has been applied in motivation research; however, few studies have examined intrinsic motivation as a moderator in gamified PE [66,67]. Sañudo et al [68] observed that gamification boosts physical activity, but their models do not account for the psychological relationship between intrinsic motivation, internal pleasure, and mental performance. To develop educational interventions that are both physically and psychologically beneficial, it is essential to understand the driving factors of gamified learning.

Research on the cognitive effects of gamified PE using AI-based feedback is seldom conducted. Although AI is becoming increasingly widespread in developing or transitional education systems, such as China's, academics have given less attention to its applications in PE. AI research in education has focused on academic learning platforms or disregarded pedagogy. Zha et al [69] study shows that AI-enhanced feedback tools may assist students in learning PE technical skills, but they do not link them to cognitive progress. In addition, there is limited knowledge about how technological tools affect motivational psychology. AI feedback in gamified PE has not been studied to determine whether it enhances self-awareness and metacognitive processing or merely motivates.

The literature on physically interactive, game-based learning environments does not address multilevel classroom dynamics, including student-teacher collaboration and instructor support. The social aspect of PE is often overlooked in empirical models that combine gamification and cognitive findings. Iglesias and Fernandez-Rio [70] conducted a comparative study. Students may be more engaged in collaborative PE. The research does not examine how these interactions increase cognition or how teacher scaffolding affects it. Comprehensive frameworks must incorporate instructional strategies, motivating factors, technological upgrades, and student interaction patterns, which affect cognitive performance.

This study examines how gamified PE, physical activity enjoyment, peer cooperation, and teacher support impact students' cognitive performance in China's unique educational system, aiming to fill these complex gaps. This study integrates pedagogical and psychological perspectives to better understand how innovative PE methods can enhance cognitive performance. The research incorporates both technology and physical exercise, unlike previous investigations. The study uses intrinsic motivation as a mediating component to enhance the theoretical understanding of internal motivational processes in gamified scenarios. It also examines how AI-based feedback moderates the effects of digital technology on learning in interactive, collaborative classrooms. The study suggests that secondary school PE programs facilitate students' cognitive and developmental progress by providing data to inform curriculum development, teacher professional development, and the effective use of educational technology. It can help Chinese and other school administrators develop balanced, tech-integrated, and student-centered PE curricula to achieve 21st-century learning goals.

Research has linked gamified PE with an intrinsic desire to exercise and improved cognitive function to form the following research hypotheses:

- H1: Gamified physical education and physical activity enjoyment have a significant positive effect on students' cognitive performance.
- H2: Peer collaboration and teacher support have a significant positive effect on students' cognitive performance.
- H3: Intrinsic motivation mediates, and AI-based feedback moderates, the relationship between gamified physical education and cognitive performance.

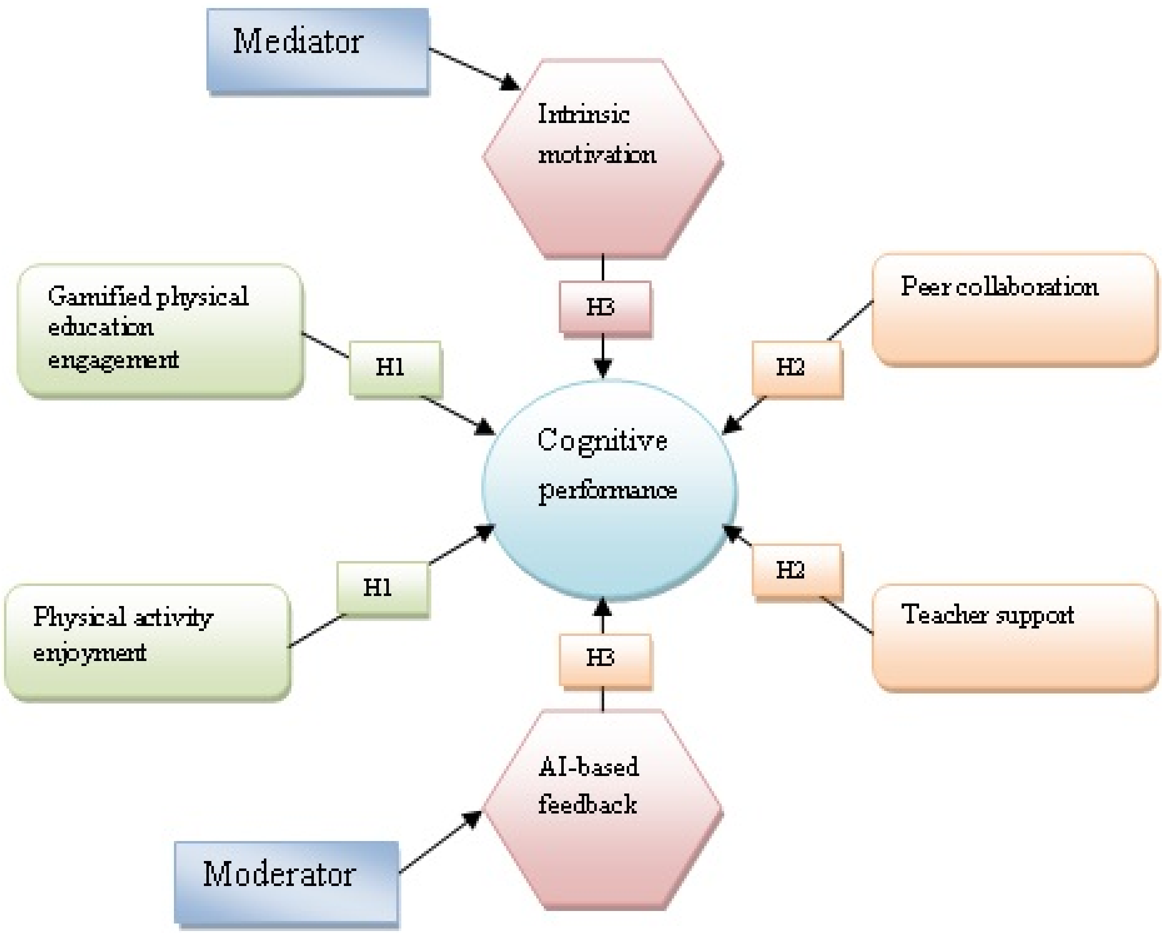
Methods

Theoretical Framework

This study is based on cognitive load theory (CLT) and self-determination theory (SDT), with contextual support from Constructivist Learning Theory. Deci and Ryan's [71] self-determination theory explains how students' sense of enjoyment in connectivity, autonomy, and competence motivates them to pursue diverse activities. SDT is helpful for this study because it shows how intrinsic motivation moderates the relationship between gamified PE and cognitive function. Gamification elements, such as points, challenges, levels, and social engagement, help PE programs fulfill students' psychological needs and encourage enthusiastic participation. As gamified PE becomes more engaging and personalized, students' cognitive engagement, memory, and problem-solving abilities are expected to improve [72]. CLT suggests that gamification and other instructional methods may increase working memory and decrease cognitive load, particularly when combined with real-time AI-based feedback [73]. Such feedback may help students focus, repair mistakes quickly, and learn more via adaptive responses by strengthening the gamified PE-cognition link. According to Constructivist Learning Theory, teachers can direct students' exploration, social interaction, and feedback, while students can also help themselves through peer collaboration and instructor support [74]. Teamwork and teacher facilitation are highly effective in promoting engagement and the application of cognitive skills in PE.

These theoretical frameworks create the study's expected variable relationships (refer to Figure 1).

Figure 1. Conceptual framework showing hypothesized relationships among explanatory variables, mediator, moderator, and dependent variable (cognitive performance) with hypotheses H1–H3. Data from 1029 high school students in Beijing, Shanghai, Chengdu, and Guangzhou, 2025. Source: author's work. AI: artificial intelligence.



The first hypothesis (H1) states that gamified PE and physical activity enjoyment will enhance students' cognitive performance. Physically active classrooms boost students' physical health, mental clarity, memory, and problem-solving skills. H2, which builds on H1, suggests that teacher and peer interaction enhances students' cognitive performance by focusing on the social and pedagogical aspects of learning. These factors make the classroom a friendly yet challenging environment for students to learn through hands-on experiences, collaborate, and receive constructive feedback. H3, the technological and psychological components, reveals that intrinsic motivation mediates the relationship between gamified PE and cognitive performance and that AI-based feedback moderates this relationship. Intrinsically motivated students are more likely to focus, continue the course, and actively engage in learning using AI feedback systems, which improves outcomes. This paradigm encompasses motivation, instruction, technology, and cognition, and when combined, these elements demonstrate how gamified PE can support learning in today's classrooms.

Population, Sample, and Research Design

The study investigated the influence of intrinsic motivation and AI-based feedback on the relationship between gamified PE and students' cognitive performance in China's unique cultural and educational context. The research includes secondary school PE students from various regions in China. The sample included urban, suburban, and rural schools from Beijing, Shanghai, Chengdu, and Guangzhou municipalities. These regions were selected because they participate in national education innovation projects and have various degrees of classroom technology integration. The study used quantitative, cross-sectional surveys and structured questionnaires to obtain participant data. The 1029 completed and valid responses from 1175 surveys issued to 10- to 18-year-olds yielded an 87.6% response rate. Based on Cochran sample size calculation approach, which accounts for a 95% CI and a 3% margin of error, a sample of desired respondents was required. Multivariate statistical analysis, including structural equation modeling (SEM), can be reliably performed with the given sample.

Demographic Profile

The study collected demographic data to understand the participants' contextual profiles and ensure diversity across crucial parameters. High, moderate, and low achievers were categorized based on their academic achievement. The demographic component of the questionnaire inquired about school type (public, private, or semigovernment), location (urban, suburban, or rural), and access to technology (high, moderate, or low). The study cross-tabulated these factors to ensure consistency among student demographics and institutional settings.

Measurement Scales

This survey included key factors from theoretical and empirical investigations (refer to Table 1). Cognitive performance was assessed using a Cognitive Functioning Scale item derived from the Woodcock-Johnson Tests. The item includes: "*My memory has improved due to gamified PE*."A cognitive development study has validated this scale [75]. The first independent variable, gamified physical education engagement, was assessed using the game-based learning engagement scale. The item was *"I feel more engaged in gamified PE*," followed by Hamari et al [76]. The second independent variable, physical activity enjoyment, was assessed using the physical activity enjoyment scale with a representative item, "*I feel energetic after physical activity*," as suggested by Kendzierski and DeCarlo [77]. The third independent variable, examined by the collaborative learning scale, was "*Team-based PE helps me learn more*," as used by Laal and Ghodsi [78] for instrumentation. Belmont et al [79] used the scale and representative items, including "*My teacher helps me reflect on progress*," to score instructor support, the fourth independent variable, on the Teacher as Social Context Questionnaire–Autonomy Support subscale. The mediating variable, intrinsic motivation, was operationalized using the intrinsic motivation inventory-interest and enjoyment subscale. One key item is "*I feel accomplished after PE*." Deci and Ryan's [71] theoretical model and psychometric validation by McAuley et al [80] provided support for this instrument. AI-based feedback was measured using a sample item from the smart education technology–enhanced feedback scale, supporting Ifenthaler et al [81] and Caspari-Sadeghi et al [82]. The study used a 5-point Likert scale for all items, with 1 indicating "strongly disagree" and 5 indicating "strongly agree."

Table 1. Measurement scales and sample items for studied variables, based on responses from 1029 high school students in 4 Chinese cities, 2025.

Constructs	Measurement scale	Sample item from questionnaire	Literature support
Cognitive performance (DV[a])	Cognitive Functioning Scale (adapted from Woodcock-Johnson Tests)	I can concentrate better during lessons after gamified PE[b].	[75]
Gamified physical education engagement (IV1[c])	Game-Based Learning Engagement Scale	I enjoy gamified PE classes.	[76]
Physical activity enjoyment (IV2[d])	Physical Activity Enjoyment Scale (PACES)	Physical activities are fun for me.	[77]
Peer collaboration (IV3[e])	Collaborative Learning Scale	I work better in teams during PE.	[78]
Teacher support (IV4[f])	Teacher as Social Context Questionnaire–Autonomy Support subscale	My PE teacher gives helpful feedback.	[79]
Intrinsic motivation (mediator)	Intrinsic Motivation Inventory (IMI): interest and enjoyment subscale	I participate in PE because I enjoy it.	[71,80]
[g]AI-based feedback (moderator)	Technology-Enhanced Feedback Scale	AI-based feedback helps me improve my physical skills.	[81,82]

[a]DV: dependent variable.

[b]PE: physical education.

[c]IV1: independent variable 1.

[d]IV2: independent variable 2.

[e]IV3: independent variable 3.

[f]IV4: independent variable 4.

[g]AI: artificial intelligence.

Pilot Study and Diagnostic Tests

A pilot study evaluated the validity, reliability, and clarity of content and timing of the research instrument with 50 students from similar backgrounds in the different groups of research participants, aged 10-18 years. After conducting reliability analysis using Cronbach , all questionnaire components had coefficients above 0.70, indicating strong internal consistency. The dimensionality of each scale was assessed using exploratory factor analysis, and construct validity, including convergent and discriminant validity, was examined using confirmatory factor analysis (CFA) with using SPSS AMOS software (IBM Corp). The scale's average variance extracted (AVE) and composite reliability (CR) values were within the acceptable range for all structures, proving its validity. Variance inflation factor multicollinearity tests verified the dataset's suitability for regression-based modeling. These tests showed that the independent variables were not highly correlated.

Reporting Standards

This study followed the APA Journal Article Reporting Standards for Studies Using Structural Equation Modeling to provide comprehensive documentation of model construction, estimation methods, model fit indices, and mediation and moderation studies [83].

SEM was used to test the study's hypotheses (H1, H2, and H3) and to estimate the direct and indirect correlations among variables for further statistical analysis. Furthermore, a moderated-mediation analysis was also performed for robust inferences. The study applied bootstrapping with 5000 resamples to establish the relevance of the indirect effects; the study accordingly performed the analysis. An interaction term was tested to assess whether AI-based feedback moderates the connection between gamified PE and cognitive performance. The relationship's intensity and direction were considered. Model fit indices, including comparative fit index (CFI), Tucker-Lewis Index (TLI), root-mean-square error of approximation (RMSEA), and standardized root-mean-square residual, were within acceptable bounds, demonstrating model robustness. The relationships were exhaustively analyzed using SPSS Statistics version 26 (IBM Corp) and AMOS version 24 (IBM Corp) for all statistical methods.

Finally, this research relies on reliable statistical methods, measurement instruments, and rigorous sampling. By incorporating both technical and psychological components into educational interventions, such as gamified PE, Chinese secondary school students can gain a multifaceted understanding of how modern pedagogical strategies impact cognitive development. The stratified sample, high response rate, and advanced SEM methodologies make the findings reliable and accessible to a large audience.

Ethical Considerations

This study was approved by the Research Ethics Committee of Kyonggi University, Suwon-si, Republic of Korea (Institutional Review Board no 12416). All procedures were conducted in accordance with institutional and national ethical guidelines. Written informed consent was obtained from all participants before data collection. For participants younger than18 years, written and oral consent was obtained from their parents or legal guardians. Trained facilitators assisted younger participants to ensure they clearly understood the survey items and participated voluntarily. All data were anonymized at the time of collection; no personal identifiers were recorded, and participants were assigned unique numeric codes for data handling and analysis. No images or other identifying materials of individual participants were collected or included in the paper. Therefore, no identifiable participant information is disclosed, and no additional consent was required for images.

Results

Stratified random sampling was used to recruit students from various schools across the 3provinces, categorized by school type and geographic location. The study initially categorized schools by geography. Each stratum's students were randomly selected. The study carefully balanced gender and grade levels for equitable sampling. PE instructors distributed questionnaires during class time with the support of administrators. Participation was voluntary, and students were given anonymity and confidentiality, which enhanced the legitimacy of the responses. School coordinators and field supervisors collaborated to gather data using standardized instruments over 2 months. Given the wide age range of participants, methodological and ethical adjustments were implemented for children in the 10- to 12-year age group, who may experience developmental difficulties in recalling intricate survey questionnaires. For this age group, the questionnaire was read aloud by trained facilitators in a structured manner. The facilitators read questions to the students individually, provided simplified explanations, and ensured that the students responded voluntarily without asking leading questions. This approach ensured the validity, accuracy, and adherence to ethical guidelines when using young respondents in psychological and educational research. The 13- to 18-year-old respondents were able to complete the questionnaire independently.

Table 2 provides the demographics of respondents to illuminate the study's sample composition and contextualize the results. The sample comprised 1029 students from various educational institutions throughout China, ensuring a broad representation of student opinions and learning conditions. Male students comprised 47.60% (n=490) of the sample, while female students comprised 52.40% (n=539). The findings are more generalizable due to a gender-diverse participant pool and relatively equal distribution. Three age groups of students were 10-12, 13-15, and 16-18 years. The students included 438 (42.60%) aged 13-15 years, 310 (30.10%) aged 10-12 years, and 281 (27.30%) aged 16-18 years. This age group encompasses a broad range of developmental periods in middle and upper elementary and secondary school, making it crucial for gamified learning and cognitive growth. The survey found that 479 (46.60%) students assessed their academic performance as moderate, 305 (29.67%) as excellent, and 245 (23.83%) as low achievers. These distributions show how students from diverse academic backgrounds use gamified PE and AI-based feedback. Public schools educated 428 (41.60%) students, private schools 311 (30.20%), and semigovernment schools 290 (28.20%). This indicates a balanced representation across institutional types, as educational systems and funding may affect the integration of gamified PE and AI. The respondents were drawn from a variety of areas, including urban (366/1029, 35.60%), suburban (360/1029, 35%), and rural (303/1029, 29.40%) locations. Geographic diversity is considered when assessing the impact of educational infrastructure and technological accessibility on cognitive performance outcomes. As a result, the study also reveals regional variations in teaching and educational resources. Finally, 400 (38.90%) students reported high access to digital tools and platforms, 451 (43.90%) reported moderate access, and 178 (17.30%) reported limited access, a significant variable given this study's focus on AI-based feedback. These results suggest that students' technological readiness may influence or moderate the effectiveness of gamified learning approaches. Given the relatively high percentages of students with moderate to high technology access, most students appeared to have participated meaningfully using digital and AI-driven teaching resources.

Table 3 provides descriptive statistics, validity indicators, and internal reliability and stability for each of the study's major latent variables, assessing the stability and internal consistency of the structural model's constructs.

Table 2. Demographic characteristics of 1029 high school student respondents across Beijing, Shanghai, Chengdu, and Guangzhou, 2025.

Demographic variable and category	Value, n (%)
Sex	
Male	490 (47.60)
Female	539 (52.40)
Age group	
10-12 years	310 (30.10)
13-15 years	438 (42.60)
16-18 years	281 (27.30)
Academic performance	
High achiever	305 (29.60)
Moderate achiever	479 (46.60)
Low achiever	245 (23.80)
School type	
Public	428 (41.60)
Private	311 (30.20)
Semigovernment	290 (28.20)
Geographic location	
Urban	366 (35.60)
Suburban	360 (35)
Rural	303 (29.40)
Technology access level	
High	400 (38.90)
Moderate	451 (43.80)
Low	178 (17.30)

Table 3. Descriptive statistics, reliability, composite reliability, and average variance extracted for all study constructs based on 1029 student responses, 2025.

Variables	Value, mean (SD)	Cronbach α	CR[a]	AVE[b]
Cognitive performance	4.18 (0.72)	0.89	0.91	0.66
Gamified PE[c] engagement	4.25 (0.68)	0.92	0.93	0.70
Physical activity enjoyment	4.12 (0.75)	0.87	0.89	0.64
Peer collaboration	4.05 (0.81)	0.86	0.88	0.61
Teacher support	4.1 (0.70)	0.91	0.92	0.69
Intrinsic motivation	4.21 (0.65)	0.90	0.91	0.68
AI[d]-based feedback	3.89 (0.79)	0.88	0.90	0.65

[a]CR: composite reliability.

[b]AVE: average variance extracted.

[c]PE: physical education.

[d]AI: artificial intelligence.

Cognitive performance, the study's primary dependent measure, had a mean score of 4.18 (SD 0.72). With modest fluctuation, participants reported strong cognitive engagement and outcomes. The scale's Cronbach of 0.89 indicated high internal consistency. The CR was 0.91, and the AVE was 0.66, both exceeding the convergent validity criteria of 0.50. Scale elements adequately explained the hidden component. Gamified PE Engagement, a major independent variable, had the highest mean score (mean 4.25, SD 0.68), reflecting positive participant views of PE gamification. Cronbach of 0.92, CR of 0.93, and

AVE of 0.70 indicate the reliability of this scale. These statistics demonstrate that the gamified engagement measurement model is statistically valid and substantively relevant due to its strong internal cohesion. Another independent variable with a high mean score of 4.12 (SD 0.75) supported favorable participant attitudes toward PE. AVE 0.64, Cronbach 0.87, and CR 0.89 all exceeded acceptable norms, indicating the construct's internal reliability. Gamified activities motivated and engaged participants in PE. Peer collaboration also had strong psychometric features, with a mean of 4.05 (SD 0.81), CR=0.88, AVE=0.61, and Cronbach α=0.86. These findings support the favorable benefits of collaborative learning on academic success and demonstrate that the instrument effectively captures the social dynamics and cooperative learning aspects of PE. Teacher support was another major independent variable with high reliability and validity. A mean of 4.10 (SD 0.70), Cronbach of 0.91, CR of 0.92, and AVE of 0.69 indicated a robust construct. These statistics highlight the role of instructors in designing, guiding, and supporting gamified education and physical learning, as well as cognitive outcomes and participant motivation. Most participants were motivated and self-driven in gamified learning contexts, with an average Intrinsic Motivation score of 4.21 (SD 0.65). This construct had an AVE of 0.68, a CR of 0.91, and a Cronbach of 0.90. Self-determination theory suggests that intrinsic motivation is crucial for academic achievement, and these values support the statistical validity of the motivation construct. Finally, AI-based feedback had a lower mean of 3.89 (SD 0.79), suggesting a wider range of responses. However, Cronbach scores of 0.88, CR 0.90, and AVE 0.65 indicate strong construct validity and internal consistency. Although experiences with AI feedback varied (possibly due to familiarity or availability), participants generally reported that it improved engagement and performance. Table 4 shows the item-level standardized loadings and metric invariance estimates.

Table 4. Item-level standardized loadings and metric invariance results for all constructs measured in 1029 high school students across 4 Chinese cities, 2025.

Variable and item	λ (standardized loading)	Cross-loading (max)	Mandarin translation ΔCFI[a]
Cognitive performance (CP)			
CP1	0.78	0.12	0.003
CP2	0.81	0.10	0.003
CP3	0.83	0.09	0.003
CP4	0.79	0.11	0.003
CP5	0.82	0.10	0.003
Gamified PE[b] engagement (GPE)			
GPE1	0.85	0.11	0.004
GPE2	0.88	0.10	0.004
GPE3	0.86	0.12	0.004
GPE4	0.84	0.09	0.004
GPE5	0.87	0.10	0.004
Physical activity enjoyment (PAE)			
PAE1	0.79	0.08	0.002
PAE2	0.81	0.07	0.002
PAE3	0.80	0.09	0.002
PAE4	0.82	0.08	0.002
PAE5	0.79	0.07	0.002
Peer collaboration (PC)			
PC1	0.75	0.10	0.003
PC2	0.78	0.11	0.003
PC3	0.77	0.09	0.003
PC4	0.76	0.08	0.003
PC5	0.79	0.10	0.003
Teacher support (TS)			
TS1	0.84	0.09	0.003
TS2	0.87	0.08	0.003
TS3	0.85	0.10	0.003
TS4	0.86	0.09	0.003
TS5	0.88	0.08	0.003
Intrinsic motivation (IM)			
IM1	0.82	0.10	0.002
IM2	0.85	0.11	0.002
IM3	0.83	0.09	0.002
IM4	0.81	0.08	0.002
IM5	0.84	0.10	0.002
AI[c]-based feedback (AIF)			
AIF1	0.80	0.09	0.003
AIF2	0.83	0.10	0.003
AIF3	0.81	0.08	0.003
AIF4	0.82	0.09	0.003

Variable and item	λ (standardized loading)	Cross-loading (max)	Mandarin translation ΔCFI[a]
AIF5	0.84	0.10	0.003

[a]CFI: comparative fit index.

[b]PE: physical education.

[c]AI: artificial intelligence.

The study looked at item-level standardized loadings (λ) and cross-loadings for all latent constructs to ensure the study instrument was reliable and valid (refer to Table 5). The standardized loadings for each item were 0.75-0.88, indicating robust construct representation, while the cross-loadings were less than 0.12, indicating discriminant validity. The translation into Mandarin was tested for metric invariance, and all constructs had ΔCFI values <0.01, which indicates that the translated items perform the same as the original instrument. Table 5 provides a comparison of original CF and higher-order CFA models, which helps determine whether the measurement model in this research is legitimate.

Table 5. Comparison of original confirmatory factor analysis (CFA) and higher-order CFA models, including model fit indices, for constructs measured in 1029 high school students in Beijing, Shanghai, Chengdu, and Guangzhou, 2025.

Model	CFI[a]	TLI[b]	RMSEA[c]	SRMR[d]	χ^2 (*df*)	AIC[e]	BIC[f]
Original CFA[g]	0.961	0.953	0.042	0.038	2.85 (532)	2180.45	2245.67
Higher-order CFA (reduced model)	0.954	0.946	0.043	0.040	2.87 (534)	2145.32	2210.58

[a]CFI: comparative fit index.

[b]TLI: Tucker-Lewis Index.

[c]RMSEA: root-mean-square error of approximation.

[d]SRMR: standardized root-mean-square residual.

[e]AIC: Akaike information criterion

[f]BIC: Bayesian information criterion

[g]CFA: confirmatory factor analysis.

In a higher-order structural factor analysis (CFA), the study combined 2latent factors: intrinsic motivation and enjoyment of physical exercise, and cognitive performance and teamwork. Having a χ^2/df ratio of 2.87 and good fit indices (CFI=0.954, TLI=0.946, RMSEA=0.043, and standardized root-mean-square residual=0.040), the higher-order model effectively fit the measurement model. The higher-order CFA has lower AIC (2145.32) and BIC (2210.58) than the original CFA, indicating that the reduced model is more parsimonious without losing theoretical or empirical validity. These results suggest that the dual-framework model (SDT + CLT) may retain construct validity and conceptual coherence with fewer latent variables.

The Fornell-Larcker criterion and the heterotrait-monotrait (HTMT) ratio of correlations were used to assess the discriminant validity of SEM. Table 6 provides their results.

Table 6. Fornell-Larcker criteria and the heterotrait-monotrait (HTMT) ratio assessing the discriminant validity of all constructs in 1029 high school students, 2025.

Constructs	Fornell-Larcker diagonal	HTMT[a] ratios (max)
Cognitive performance	0.81	0.82
Gamified PE[b]	0.84	0.79
PE[b] enjoyment	0.80	0.76
Peer collaboration	0.78	0.74
Teacher support	0.83	0.77
Motivation	0.82	0.81
AI[c] feedback	0.81	0.80

[a]HTMT: heterotrait-monotrait.

[b]PE: physical education.

[c]AI: artificial intelligence.

The diagonal line shows the square root of each construct's AVE compared to the interconstruct correlations, using Fornell-Larcker criteria. This method demonstrates discriminant validity when the square root of a concept's AVE is greater than its highest correlation with any other construct. Cognitive performance's Fornell-Larcker diagonal value of 0.81 is greater than its connections with gamified PE, motivation, and AI feedback. AI feedback (0.81), gamified PE (0.84), PE enjoyment (0.80), peer collaboration (0.78), teacher support (0.83), motivation (0.82), and AI Support (0.83) all have top diagonal values, indicating that each construct's items are more strongly related to their latent variable than to any other. The HTMT ratio was used to enhance this investigation. PLS-SEM applies HTMT for a more sensitive and reliable discriminant validity test. Discriminant validity is good when HTMT values are <0.85. All HTMT scores in the results were between 0.74 and 0.82. For cognitive performance, none of the constructs has an HTMT ratio >0.82, which is sufficient. AI feedback (0.80), gamified PE (0.79), PE enjoyment (0.76), peer collaboration (0.74), teacher support (0.77), motivation (0.81), and AI support (0.79) all fall below the threshold, supporting the premise that each latent concept is empirically distinct. Table 7 shows the estimates of Harman single-factor and latent common method variance.

Table 7. Harman single-factor and latent common method variance (CMV) tests to evaluate potential survey bias in 1029 high school student responses, 2025.

Test type	No of factors extracted	Variance explained by first factor (%)	Total variance explained (%)	ΔCFI[a] (with CMV[b] factor)	$\Delta\chi^{2c}$ (*df*)	*P* value	Conclusion
Harman single-factor test	7	28.6	72.4	—	—	—	No CMV concern (first factor<40%)
Latent common method factor test	7 + CMV	26.9	74.1	0.006	21.38 (1)	.09	No significant CMV effect (ΔCFI<0.01)

[a]CFI: common method variance.

[b]CMV: common method variance.

[c]$\Delta\chi^2$: change in chi-square value.

Harman single-factor test and the latent CMV factor test were used to assess for self-reported data-related CMV. The unrotated factor analysis found 7 variables, although the first accounted for 28.6% of the variance (vs 40%). This indicates that no factor dominated the item-level covariance structure. This finding was confirmed by adding a latent CMV component to the measurement model. The model fit comparison showed no significant reduction in fit, even after controlling for method effects (ΔCFI=0.006 and $\Delta\chi^2_1$=21.38; *P*=.09). These findings strongly suggest that common method bias does not threaten the study's validity and that discriminant integrity of latent constructs is preserved.

Table 8 provides the SEM findings, which strongly support the research's theoretical approach. There is a positive relationship between gamified PE and students' cognitive performance. Gamified PE may enhance students' concentration, memory, and information-processing skills [84]. According to the engagement-learning paradigm, students learn meaningfully when they are emotionally, behaviorally, and cognitively involved [85]. Game elements, such as goal-setting, fast feedback, and reward systems, in PE sessions, may help children learn more effectively and improve cognitively [86]. Chaiyarat [87] and Aibar-Almazán et al [88] reported that gamification can make the classroom more dynamic and engaging, thereby enhancing students' problem-solving and critical-thinking skills.

Table 8. Structural equation modeling (SEM) path estimates showing relationships among studied factors with standardized beta coefficients and significance levels, based on 1029 students, 2025.

Path	Estimates (β)	SE	*t* statistic	*P* value
Gamified PE[a] → cognitive performance	.34[b]	0.05	6.80	<.001
Enjoyment → cognitive performance	.28[b]	0.06	4.67	<.001
Collaboration → cognitive performance	.26[b]	0.07	3.71	<.001
Teacher support → cognitive performance	.31[b]	0.06	5.17	<.001

[a]PE: physical education.

[b]*P*<.001 (1% significance level).

Teacher support is the second strongest predictor of students' cognitive performance. This suggests that instructors must understand how to positively impact students' cognitive growth through effective lesson preparation, positive reinforcement, and constructive feedback. Gamified and AI-feedback classrooms require teacher facilitation [89]. Teachers should motivate students by establishing a psychologically safe classroom, helping them grasp and implement challenging feedback, and scaffolding their learning to support their growth and development [90]. Cha et al [91] found that students feel more confident, engaged, and cognitively involved in classroom tasks, especially in active and digitally mediated settings, when they view their teachers as accessible, attentive, and helpful.

There is a positive relationship between physical activity enjoyment and students' cognitive performance, which suggests that students who enjoy PE may focus better, experience lower cognitive stress, and feel emotionally good, which helps them learn. According to SDT, individuals are more engaged and learn more when they have a personal interest in the result [92]. Chen [93] found that children who enjoy physical activities are more willing to participate and better able to harness the psychological benefits of exercise, leading to improved classroom concentration, memory, and performance. Enjoyment enhances cognition in gamified environments by reducing performance anxiety and boosting self-confidence.

Peer collaboration and cognitive performance were significantly positively associated. Collaborating with peers during gamified PE classes boosts cognitive development. The focus on idea-sharing, collaborative problem-solving, and social and emotional support in peer learning enhances understanding and mental agility. This result is supported by Vygotsky's sociocultural theory, which emphasizes the role of social interaction in internalizing information [94]. Qi and Derakhshan [95] found that physically active educational activities help students develop social and cognitive skills for academic success. Collaboration, negotiation, and peer feedback improve metacognition and learning.

The study analyzed cluster effects in Beijing, Shanghai, Chengdu, and Guangzhou using intraclass correlation coefficients (ρ) and design effects for each latent construct. The intraclass correlation coefficients ranged from 0.02 to 0.06, resulting in design effects of 7.18 to 19.90 with an average cluster size of 310 respondents per city (Table 9). After achieving the threshold (ρ>0.05), a multilevel structural equation model with random intercepts was used to review cognitive performance and AI-based feedback. There were minor differences in explained variance (ΔR^2= +0.02 to +0.05) and standardized path coefficients (Δβ<.02) between single-level and multilevel estimations. The findings reveal that provincial clustering did not significantly affect gamified PE engagement, intrinsic motivation, and cognitive performance. Therefore, the estimation technique does not influence structural model robustness.

Table 9. Cluster diagnostics and multilevel structural equation modeling (SEM) robustness checks verifying the stability of results across school clusters in 1029 high school students, 2025.

Construct	Number of clusters (cities)	Average cluster size ($\bar{n}$)	ICC[a] (ρ)	Design effect=1+($\bar{n}$−1)ρ	Single-level β (gamified PE → cognitive performance)	Multilevel β (random intercept model)	Δβ[b] (bias)	ΔR^2 (change in explained variance)
Cognitive performance (CP)	4 (Beijing, Shanghai, Chengdu, and Guangzhou)	310	0.05	16.45	.45	.43	−.02	0.04
Gamified PE[c] engagement	4	310	0.04	13.36	.34	.33	−.01	0.03
Intrinsic motivation	4	310	0.03	10.27	.21 (indirect κ^2=0.16)	.20	−.01	0.02
AI[d]-based feedback	4	310	0.06	19.9	.18 (interaction β)	.17	−.01	0.05
Teacher support	4	310	0.02	7.18	.31	.30	−.01	0.02
Peer collaboration	4	310	0.03	10.27	.26	.25	−.01	0.02

[a]ICC: intraclass correlation coefficient.

[b]Δβ: change in bias.

[c]PE: physical education.

[d]AI: artificial intelligence.

Table 10 assessed the impact of AI-based feedback on the model using effect-size indices (κ^2 and PM), conditional effects at ±1 SD, and ΔR^2. The mediation study revealed a medium-to-large effect size (κ^2=0.19) and PM=0.36, suggesting that intrinsic motivation indirectly accounts for36% of the connection. Intrinsic motivation substantially influenced the link between gamified PE and cognitive performance (β=.23; *P*<.001). A moderated analysis indicated that AI-based feedback increased the positive correlation between gamified PE and cognitive performance (β=.17; *P*<.001). The correlation between AI feedback and gamified learning gains increased with complexity, with β=.13 at low levels (–1 SD), β=.20 at medium, and β=.28 at high levels (+1 SD). The inclusion of the interaction term raised the model's explanatory power from R^2=0.47 to R^2=0.55, resulting in a ΔR^2 of 0.08, indicating that AI-based feedback explained an additional 8% of the variance in students' cognitive performance.

Table 10. Moderation-mediation analysis showing effect sizes, conditional effects (±1 SD of artificial intelligence [AI]–based feedback), and ΔR² for relationships between gamified physical education (PE), intrinsic motivation, and cognitive performance in 1029 students, 2025.

Path	Estimate (β)	SE	*t* value	*P* value	LLCI[a]	ULCI[b]	κ^2[c]	PM[d]	Conditional effect (–1 SD)[e]	Conditional effect (mean)[e]	Conditional effect (+1 SD)[e]	R^2	ΔR^2[f]
Mediation effect: gamified PE →intrinsic motivation →cognitive performance	.23[g]	0.04	5.75	<.001	0.15	0.32	0.19	.36	—	—	—	0.47	—
Moderation effect: gamified PE × AI feedback →cognitive performance	.17[g]	0.03	5.67	<.001	0.10	0.24	—	—	0.13	0.20	0.28	0.55	+0.08
Total effect: gamified PE →cognitive performance (including mediation and moderation)	.45[g]	0.05	9.00	<.001	0.35	0.55	—	—	—	—	—	—	—

[a]LLCI: lower limit CI.

[b]ULCI: upper limit CI.

[c]κ^2: standardized indirect effect size.

[d]PM: proportion mediated.

[e]Conditional effects: changes in the slope of gamified PE → cognitive performance at low (–1 SD), mean, and high (+1 SD) levels of artificial intelligence (AI)–based feedback.

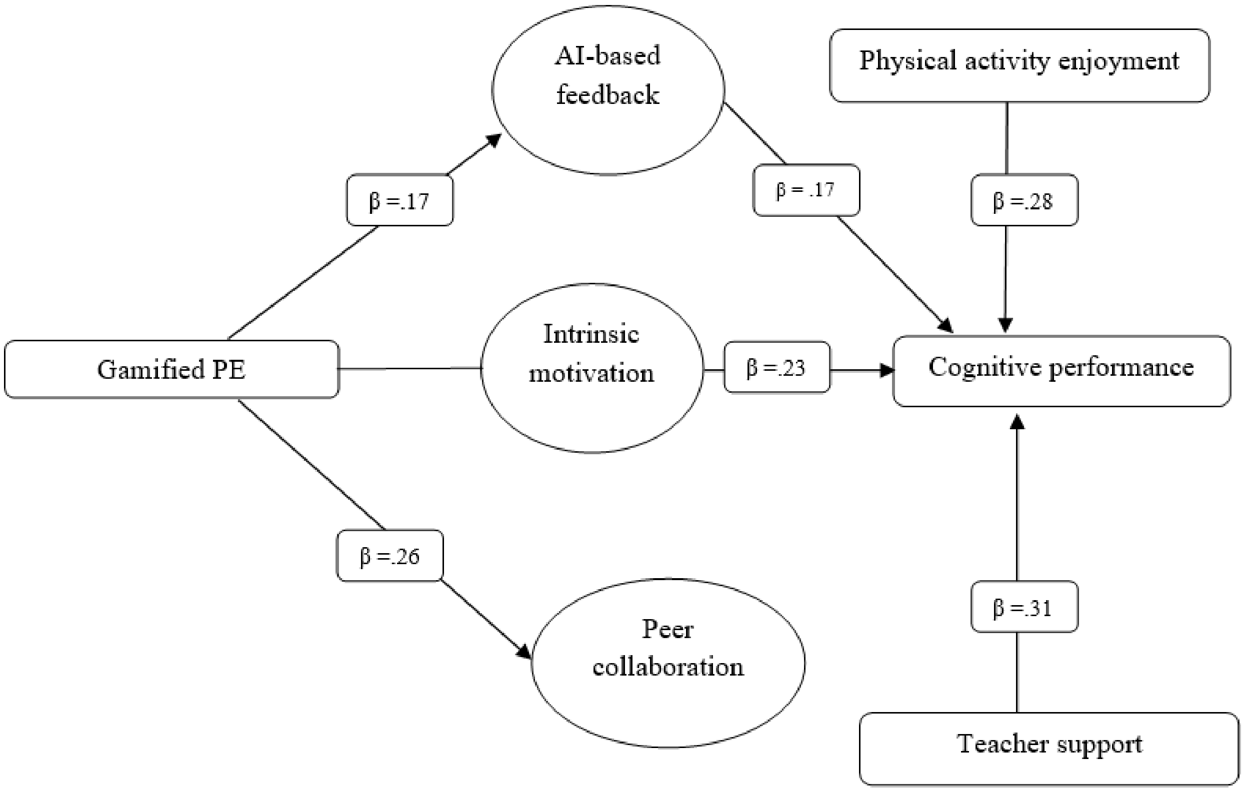
[f]ΔR^2: 0.08, indicating an 8% increase in explained variance after counting the moderation term.

[g]*indicates a 1% significance level.

The quantity of AI feedback in gamified PE sessions affects student engagement and cognitive performance. Gamified PE improves cognitive function, and AI-based feedback enhances this impact. Children who participate in gamified PE and receive continuous, real-time AI-powered feedback are more likely to experience enhanced cognitive outcomes, including improved memory recall, attention, problem-solving, and mental engagement [96]. Feedback intervention theory suggests that timely and personalized feedback helps learners concentrate on task-related goals, self-regulate, and engage cognitively [97]. It may be challenging to obtain adaptive, data-driven insights in traditional PE settings, but AI-powered feedback is immediate, objective, and tailored. These systems, integrated into a gamified framework, make learning more engaging and challenging for students, providing explicit guidance on how to improve, which in turn fosters deeper thinking. According to Suresh Babu and Dhakshina Moorthy [98], gamification alone can boost student engagement. However, intelligent feedback systems can amplify these effects on cognition by influencing learner behavior and performance in real-time. The study's findings support the use of technology-enhanced, customized learning environments in modern school design. Gamified education uses AI feedback as a cognitive support system to help students recall and apply what they have learned via hands-on, interactive activities [99]. CLT suggests that real-time AI feedback, which optimizes task difficulty and decreases ambiguity, helps learners focus on important cognitive activities. The moderating impact also affects curriculum and educational policy, suggesting that complex feedback systems are needed for gamified physical instruction to maximize cognitive outcomes. For gamified learning to be most effective, schools and instructors should use or invest in AI-powered solutions that tailor insights and feedback to each student's unique profile and cognitive capabilities.

Gamified PE improves cognitive function, and intrinsic motivation is a crucial psychological factor. Gamified PE enhances intrinsic motivation, leading to improved cognitive performance. The indirect effect explains most of the variance in cognitive performance, confirming that gamification's structure and qualities are significant, but what counts most are learners' psychological moods. Self-determination theory emphasizes relatedness, competence, and autonomy as components of intrinsic motivation, which is reflected in this mediation effect. Gamified learning environments enhance intrinsic interest and satisfaction by fostering autonomy through choice, competence through manageable tasks, and relatedness through peer collaboration [53]. Students who are genuinely motivated to study are more likely to use metacognitive skills, pay attention, and learn more deeply. Shalgimbekova et al [100] suggest that motivation moderates the link between instructional design and learning outcomes. This result highlights the role of motivation in mediating the pedagogical efficacy of innovative teaching methods. Even if gamified PE is fun and structured, internalizing values and desires drives cognitive growth. This study supports the idea that the motivating processes of instructional inputs are as essential as the inputs themselves in cognitive performance [101]. Gamification sets the scene, but intrinsic motivation propels cognitive functioning. This mediation strengthens Hypothesis H3, which states that intrinsic motivation mediates the relationship between gamified PE and cognitive performance. Gamification alone is ineffective; however, students' natural incentive to participate and accomplish tasks substantially enhances outcomes, as evidenced by the considerable indirect path. This means that instructional designers, policymakers, and instructors should support children's intrinsic drive to learn and incorporate engaging aspects into PE interventions to enhance academic cognition. Figure 2 shows the structural model with standardized path coefficients.

Figure 2. Structural model displaying standardized path coefficients for all hypothesized relationships among studied factors in 1029 high school students, 2025. Source: Author's estimate.AI: artificial intelligence; PE: physical education.



The significant first-order factor loadings (λ=0.65-0.88; *P*<.001) indicate high correlations between observable indicators and their latent components (refer to Table 11). The substantial second-order factor loadings (λ=0.71-0.84; *P*<.001) suggest that SDT and CLT collaborate to create a more sophisticated integrative model. The complete model suited the specified conceptual framework well, with CFI=0.957, TLI=0.949, and RMSEA=0.041.

Table 11. Structural equation modeling (SEM) results for the dual-framework model with standardized path coefficients in 1029 students, 2025.

Analysis	Statistic	*P* value
Second-order CFA[a]	Factor loadings: λ=0.65-0.88 (1st-order) and λ=0.71-0.84 (2nd-order)	<.001
Model fit indices	CFI[b]=0.957, TLI[c]=0.949, and RMSEA[d]=0.041	—[e]
Covariance	Covariance=0.42	<.001
Cross-loading check	Max cross-loading=0.25	—[e]
Latent interaction test	$\Delta\chi^2$=4.12[f]	.04

[a]CFA: confirmatory factor analysis.
[b]CFI: comparative fit index.
[c]TLI: Tucker-Lewis Index.
[d]RMSEA: root-mean-square error of approximation.
[e]Not available.
[f]Indicates a 5% significance level.

Further, a significant association (covariance=0.42; *P*<.001) was found between intrinsic motivation and AI-based feedback, using a covariance approach. While preserving discriminant validity, this study supports the significant link between these factors. The measurement model was tested, and cross-loadings indicated that cognitive performance, intrinsic motivation, and AI-based feedback are empirically independent, supporting the structural coherence of the dual-framework SEM.

To determine whether AI-based feedback has context-specific effects, the study conducted multigroup moderation tests by level of technology access (high, moderate, and low) and location (urban vs rural; refer to Table 12).

Table 12. Multigroup moderation analysis examining differences in study relationships by urban vs rural location and technology access in 1029 high school students across 4 cities, 2025.

Path	Group	Sample size (n)	β estimate	SE	*t* value	*P* value	ΔR^2[a]	f²[b]
Gamified PE[c] → cognitive performance	Urban	366	.35[d]	0.06	5.83	<.001	0.05	0.16
Gamified PE → cognitive performance	Rural	303	.31[d]	0.07	4.43	<.001	0.05	0.15
Gamified PE → cognitive performance	High-tech access	400	.37[d]	0.05	7	<.001	0.07	0.18
Gamified PE → Cognitive Performance	Moderate tech access	451	.34[d]	.05	6.8	<.001	0.05	0.16
Gamified PE → cognitive performance	Low-tech access	178	.28[d]	0.06	4.67	<.001	0.04	0.14

[a]ΔR^2: *P*<.05.
[b]f²: *P*<.05.
[c]PE: *P*<.05.
[d]Indicates a 1% significance level.

Table 12 shows a substantial influence of gamified PE on cognitive performance across all groups (β=.28-.37; *P*<.001). Impacts were higher for urban students and those with high technological access (β=.35 and .37, respectively) and lower for rural students and those with low access (β=.31 and .28, respectively). Moderation had medium to high impacts, with ΔR^2 values ranging from 0.04 to 0.07 and f² effect sizes from 0.14 to 0.18. These findings suggest that gamified PE interventions based on AI are generally effective, with urban students and those with greater technological access benefiting more cognitively.

Discussion

Principal Findings

This study examined how intrinsic motivation and AI-based feedback affect high school students' cognitive performance after gamified PE. The first objective was to explore the effects of gamified PE and teacher support on students' cognitive performance; statistically significant benefits were shown, confirming Hypothesis 1. The second goal was to assess how physical exercise enjoyment and peer collaboration affect cognitive performance, which was confirmed by the strong positive connections between these factors and cognitive performance, supporting Hypothesis 2. The findings supported the third aim, that is, determining whether gamified PE influences cognitive performance and whether intrinsic motivation mediated this impact. The positive effect of gamified PE on cognitive performance was partially mediated by intrinsic motivation, but AI-based feedback enhanced it, supporting Hypothesis 3. Gamified PE improves cognitive performance through motivational and technical processes, in line with the research's goals and theoretical framework.

Interpretation and Implications

Gamified PE improves cognitive function, suggesting that schools can create engaging, dynamic, and technologically advanced learning environments. Gamification boosts cognitive functioning by focusing students, simplifying problem-solving, and stimulating active involvement. Well-designed, gamified PE interventions may achieve these aims, supporting the view of Barz et al [102] that game-based learning may boost motivation and cognition. The SDT concept was reinforced by intrinsic motivation, which asserts that students who feel autonomous, competent, and socially engaged are more likely to participate in meaningful activities and achieve higher-order cognitive achievements. This suggests that gamified PE improves cognition and intrinsic motivation, which sustains engagement and learning [103]. Student involvement and academic success are highest in PE programs that incorporate enjoyable and independent activities. Social and instructional factors, including a friendly teacher, exercise, and teamwork, promote learning. Game-based and AI-supported learning benefit from human facilitation, although instructors' guidance, encouragement, and scaffolding are still needed. Teamwork promotes collaboration, communication, and executive functioning, which may explain its cognitive advantages. These findings support more comprehensive theoretical frameworks that promote student-centered, socially engaged, and holistic learning environments [104]. AI-based feedback moderates the effect of adaptive, real-time recommendations on learning. AI-assisted strategy correction, refinement, and progressive advancement, delivered through rapid, tailored feedback, boosts cognitive performance beyond the reach of gamification. Recent AI-assisted learning experiments have shown potential for cognitive optimization and personalized skill improvement [105].

Comparison With Existing Literature

This study builds on previous classroom gamification research. Gamification and AI feedback in PE can enhance cognitive performance, contrary to previous research on academic participants [106,107]. While gamification has been shown to improve motivation, little research has linked it to cognitive outcomes in PE [108,109]. These findings suggest that technology-driven feedback and compelling design may boost students' cognitive performance. The study contributes to AI in education research by confirming previous classroom-based studies [110,111] that real-time adaptive feedback increases learning. This study combines SDT and CLT to advance theory. Gamified experiences increase learning via intrinsic motivation and cognitive load. These results provide a foundation for future research on integrated, technologically advanced, motivation-based educational interventions.

Practical Implications

The findings suggest that AI-based feedback mechanisms could be beneficial to educational institutions, particularly in technologically advanced cities. To achieve this, AI-powered educational systems must be funded. These systems should have adaptive capabilities and age-appropriate user interfaces to provide children with timely, individualized feedback on their PE, behavioral engagement, and cognitive attention. Policy cooperation among Chinese education ministries, IT businesses, and AI research institutes may accelerate the development of educational systems tailored to the nation. To ensure that technology enhances human teaching, a national AI-in-education framework must guide deployment, data ethics, privacy protection, and teacher capacity-building. Pilot studies in schools with suitable digital infrastructure may examine the effects of AI-enhanced, gamified learning on academic achievement. Since public and semigovernment institutions had larger performance disparities, the findings imply a shift toward socially and motivationally enhanced learning. Policies should prioritize training PE instructors to create welcoming, inclusive, and inspirational learning environments. To stimulate cognitive development through intrinsic motivation and peer collaboration, schools must prioritize students' mental health, incorporate team-building activities, and use inclusive instructional techniques that cater to their unique social identities and needs. Teacher performance measurements should incorporate student-centered learning, emotional support, and inclusive engagement as part of education reforms. Physical exercise, reflection, feedback, and cognitive challenge must be balanced in gamified learning settings that promote collaboration and active learning.

Limitations

This study has made many essential contributions. However, there are some limitations. Although the research used a large and diverse sample from 4 metropolitan cities, its cross-sectional design limits causal inferences. SEM and bootstrapping provide more compelling findings; however, experimental or longitudinal designs are necessary to determine the long-term impact of gamified PE and AI feedback on cognitive function. Second, self-report assessments may be biased by social desirability or cognitive misunderstanding, even when administered under supervision to children aged 10-12 years. Future studies could triangulate findings and improve measurement validity by using multi-informant data, such as behavioral observations or teacher ratings. The generalizability of the results is another issue. The sample included students from urban and suburban public, private, and semigovernment schools, but not rural or low-tech institutions. This raises the question of how AI-based gamified systems can effectively serve diverse socioeconomic and geographical conditions. Future research should examine how infrastructural variations affect such efforts and the digital divide. Although this study focused on cognitive function, it did not examine emotional regulation, academic resilience, or physical health. Future multidimensional research should examine how gamified and AI-supported instructional techniques affect broader student capacities. This study focused on AI feedback as a moderator and intrinsic motivation as a mediator. Other factors, such as self-efficacy, goal orientation, and support from family and friends, influence students' attitudes toward gamified learning. Future studies should incorporate additional mediators and moderators to gain a deeper understanding of the complex interplay among human, systemic, and technological factors. In conclusion, qualitative research methods, such as student interviews or classroom ethnographies, can complement quantitative approaches by providing contextual insights into learners' experiences and preferences. By addressing these limitations, future research may enhance the theoretical and practical understanding of how gamified, technology-enhanced education can improve student learning and growth.

Conclusions

The study concludes that Gamified PE with intrinsic motivation and AI-based feedback improves high school students' cognitive performance. These findings demonstrate that technology-based student-centered PE programs improve engagement, motivation, and cognition. Gamification and AI in education may improve students' overall development and academic and social results outside of the classroom. Future research should examine the long-term impacts, demographic factors, and technological developments to enhance the cognitive and motivational benefits of school-based treatments. To capitalize on the full potential of gamified PE for cognitive development, policymakers in education must institutionalize gamified pedagogy into the national PE curriculum. Schools should update their PE curriculum to highlight the connection between physical health and emotional and intellectual development. The curriculum should incorporate more structured, game-based learning methods that enhance cognitive abilities. In PE, goal-setting, point-scoring, and challenge-based learning help students concentrate, recall, and solve issues. The Ministry of Education, along with state and regional education bureaus, should educate PE teachers in gamified teaching methods to meet pedagogical criteria, engage students, and measure academic achievement.

This study advances the literature by introducing a technology-driven model of PE that combines gamification with AI-based feedback to enhance students' cognitive abilities. This study demonstrates the synergistic effects of gamification, intrinsic motivation, and AI-supported feedback on cognitive development, in contrast to previous research that has primarily focused on conventional PE or singular teaching methodologies. This study contributes to the existing literature on digital and AI-enhanced education by using a substantial, multicity sample and SEM to demonstrate the validity of these connections. The findings have substantial implications for educational institutions, educators, and policymakers, suggesting that interactive, AI-driven PE programs can effectively enhance student engagement, learning processes, and cognitive performance in genuine educational environments.

Acknowledgments

No generative artificial intelligence tools were used in the writing, editing, data analysis, or preparation of this paper.

Data Availability

The dataset generated and analyzed during this study are not publicly available due to privacy protections for minors and institutional restrictions. Deidentified data may be made available upon reasonable request to the corresponding author, subject to ethical approval.

Funding

No external financial support or grants were received for this work.

Authors' Contributions

JZ contributed to the study's conceptualization, methodology development, data collection, data curation, formal analysis, and drafting of the original paper. SSO supported the conceptualization of the research, assisted with data collection, and contributed to critical review and paper editing. YX supervised the overall study, provided guidance throughout the research process, contributed to study design and interpretation of findings, participated in paper revision, and managed project administration as the corresponding author.

Conflicts of Interest

None declared.

References

1. Guo X, Chen S, Guo Y. Advancing multimodal teaching: a bibliometric and content analysis of trends, influences, and future directions. Humanit Soc Sci Commun 2024;11(1):1-22. [doi: 10.1057/s41599-024-04254-0]
2. Triola MM, Rodman A. . Integrating generative artificial intelligence into medical education: curriculum, policy, and governance strategies. Academic Medicine 2025;100(4):413-418. [doi: 10.1097/ACM.0000000000005963]
3. Mallek F, Mazhar T, Faisal Abbas Shah S, Ghadi YY, Hamam H. A review on cultivating effective learning: synthesizing educational theories and virtual reality for enhanced educational experiences. PeerJ Comput Sci 2024;10:e2000 [FREE Full text] [doi: 10.7717/peerj-cs.2000] [Medline: 38855256]

4. Latino F, Romano G, Tafuri F. Physical education teacher's continuing professional development affects the physiological and cognitive well-being of school-age children. Education Sciences 2024;14(11):1199. [doi: 10.3390/educsci14111199]
5. Christopoulos A, Mystakidis S, Stylios C, Tsoulos IG. Unplugged gamification in education: developing computational thinking skills through embodied gameplay. J Educ Res 2025;119(1):77-90 [FREE Full text] [doi: 10.1080/00220671.2025.2517266]
6. Martín-Rodríguez A, Gostian-Ropotin LA, Beltrán-Velasco AI, Belando-Pedreño N, Simón JA, López-Mora C, et al. Sporting mind: the interplay of physical activity and psychological health. Sports (Basel) 2024;12(1):37 [FREE Full text] [doi: 10.3390/sports12010037] [Medline: 38275986]
7. Shah P, Gadekar NA, Taylor SS. Escape from passive learning: harnessing gamified digital educational escape rooms for immersive marketing education. J Mark Educ 2024;47(3):211-231 [FREE Full text] [doi: 10.1177/02734753241306093]
8. Lampropoulos G, Sidiropoulos A. Impact of gamification on students' learning outcomes and academic performance: a longitudinal study comparing online, traditional, and gamified learning. Education Sciences 2024;14(4):367. [doi: 10.3390/educsci14040367]
9. Kim J, Lindqvist AK, Castelli DM. Feasibility of utilizing gamified learning as a motivational strategy for promoting physical activity participation and healthy eating among college students. J Am Coll Health 2025;73(6):2651-2659. [doi: 10.1080/07448481.2024.2325929] [Medline: 38546660]
10. Guamanga MH, Saiz C, Rivas SF, Almeida LS. Analysis of the contribution of critical thinking and psychological well-being to academic performance. Front. Educ 2024;9:1423441. [doi: 10.3389/feduc.2024.1423441]
11. Zhang Y. Impact of arts activities on psychological well-being: emotional intelligence as mediator and perceived stress as moderator. Acta Psychol (Amst) 2025;254:104865 [FREE Full text] [doi: 10.1016/j.actpsy.2025.104865] [Medline: 40049081]
12. Sanaeifar F, Pourranjbar S, Pourranjbar M, Ramezani S, Mehr SR, Wadan AS, et al. Beneficial effects of physical exercise on cognitive-behavioral impairments and brain-derived neurotrophic factor alteration in the limbic system induced by neurodegeneration. Exp Gerontol 2024;195:112539 [FREE Full text] [doi: 10.1016/j.exger.2024.112539] [Medline: 39116955]
13. Martín-Rodríguez A, Madrigal-Cerezo R. Technology-enhanced pedagogy in physical education: bridging engagement, learning, and lifelong activity. Education Sciences 2025;15(4):409. [doi: 10.3390/educsci15040409]
14. Li Q, Li L, He X, Wang H. Exploring adolescent moderate-to-vigorous physical activity in China: mediating roles of school climate, perceived barriers, and physical education satisfaction. Risk Manag Healthc Policy 2024;17:3125-3136 [FREE Full text] [doi: 10.2147/RMHP.S497472] [Medline: 39687748]
15. Zhang J. Shaping the future: analysing early childhood education policy reforms and development in China. Euro J of Education 2025 Feb 21;60(1):e70055. [doi: 10.1111/ejed.70055]
16. Zhou S. Gamifying language education: the impact of digital game-based learning on Chinese EFL learners. Humanit Soc Sci Commun 2024;11(1):1-14. [doi: 10.1057/s41599-024-04073-3]
17. Ryan RM, Deci EL. Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. Contemp Educ Psychol 2020;61:101860. [doi: 10.1016/j.cedpsych.2020.101860]
18. Sotos-Martínez VJ, Tortosa-Martínez J, Baena-Morales S, Ferriz-Valero A. It's game time: improving basic psychological needs and promoting positive behaviours through gamification in physical education. Eur Phys Educ Rev 2023;30(3):435-457. [doi: 10.1177/1356336x231217404]
19. Xiao Y, Hew KFT. Intangible rewards versus tangible rewards in gamified online learning: which promotes student intrinsic motivation, behavioural engagement, cognitive engagement and learning performance? Brit J Educational Tech 2023;55(1):297-317. [doi: 10.1111/bjet.13361]
20. Hong H, Dai L, Zheng X. Advances in wearable sensors for learning analytics: trends, challenges, and prospects. Sensors (Basel) 2025;25(9):2714. [doi: 10.3390/s25092714] [Medline: 40363153]
21. Wang Y. Report on smart education in China. In: Zhuang R, editor. Smart Education in China and Central & Eastern European Countries. Singapore: Springer; 2023.
22. Cao F, Xiang M, Chen K, Lei M. Intelligent physical education teaching tracking system based on multimedia data analysis and artificial intelligence. Mob Info Sys 2022;2022:7666615. [doi: 10.1155/2022/7666615]
23. Li N. Fostering creative thinking through gamified intercultural communication: innovative pedagogical strategies in higher education English programs. Cogent Education 2025;12(1):2489271. [doi: 10.1080/2331186x.2025.2489271]
24. Gómez-Carrasco CJ, Monteagudo-Fernández J, Moreno-Vera JR, Sainz-Gómez M. Evaluation of a gamification and flipped-classroom program used in teacher training: perception of learning and outcome. PLoS One 2020;15(7):e0236083 [FREE Full text] [doi: 10.1371/journal.pone.0236083] [Medline: 32673373]
25. Zhu T, Gu M. The application and implications of science and technology innovation in the management of education for Chinese students studying abroad in China in the epidemic era. PLoS One 2024;19(8):e0306785. [doi: 10.1371/journal.pone.0306785] [Medline: 39150972]
26. Zhao B, Zhang Y. The impact of a personality trait rooted in Chinese confucian culture-Junzi personality-on loneliness and subjective well-being among Chinese individuals: the mediating role of self-reflection and self-control. BMC Psychol 2025;13(1):137. [doi: 10.1186/s40359-025-02488-4] [Medline: 39972519]

27. Wong D, Liu H, Meng-Lewis Y, Sun Y, Zhang Y. Gamified money: exploring the effectiveness of gamification in mobile payment adoption among the silver generation in China. ITP 2021;35(1):281-315. [doi: 10.1108/itp-09-2019-0456]
28. Ishaq K, Mat Zin NA, Rosdi F, Jehanghir M, Ishaq S, Abid A. Mobile-assisted and gamification-based language learning: a systematic literature review. PeerJ Comput Sci 2021;7:e496 [FREE Full text] [doi: 10.7717/peerj-cs.496] [Medline: 34084920]
29. Taşkın N, Kılıç Çakmak E. Effects of gamification on behavioral and cognitive engagement of students in the online learning environment. Int J Hum-Comput Interact 2022;39(17):3334-3345. [doi: 10.1080/10447318.2022.2096190]
30. Arufe-Giráldez V, Sanmiguel-Rodríguez A, Ramos-Álvarez O, Navarro-Patón R. Gamification in physical education: a systematic review. Educ Sci 2022;12(8):540. [doi: 10.3390/educsci12080540]
31. Wei Y, Wang L, Tang Y, Deng H, Su J, Li Q. Enhancing young children's executive function through physical activities: a three-level meta-analysis. Ment Health Phys Act 2024;26:100592. [doi: 10.1016/j.mhpa.2024.100592]
32. Camacho-Sánchez R, Manzano-León A, Rodríguez-Ferrer JM, Serna J, Lavega-Burgués P. Game-based learning and gamification in physical education: a systematic review. Educ Sci 2023;13(2):183. [doi: 10.3390/educsci13020183]
33. Christopoulos A, Mystakidis S. Gamification in education. Encycl 2023;3(4):1223-1243. [doi: 10.3390/encyclopedia3040089]
34. Hsu CY, Wu TT. Application of business simulation games in flipped classrooms to facilitate student engagement and higher-order thinking skills for sustainable learning practices. Sustainability 2023;15(24):16867. [doi: 10.3390/su152416867]
35. Du Q. How artificially intelligent conversational agents influence EFL learners'self-regulated learning and retention. Educ Inf Technol 2025;30(15):21635-21701 [FREE Full text] [doi: 10.1007/s10639-025-13602-9]
36. Eng CM, Tsegai-Moore A, Fisher AV. Incorporating evidence-based gamification and machine learning to assess preschool executive function: a feasibility study. Brain Sci 2024;14(5):451 [FREE Full text] [doi: 10.3390/brainsci14050451] [Medline: 38790430]
37. Zang W, Zhu J, Xiao N, Fang M, Li D, Li H, et al. Effects of aerobic exercise on children's executive function and academic performance: a systematic review and meta-analysis. Heliyon 2024;10(7):e28633 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e28633] [Medline: 38586404]
38. Ahmad F, Ahmed Z, Obaid I, Shaheen M, Rahman H, Singh G, et al. Motivation, engagement, and performance in two distinct modes of brain games: a mixed methods study in children. International Journal of Human–Computer Interaction 2024;41(15):9652-9666 [FREE Full text] [doi: 10.1080/10447318.2024.2426915]
39. Mo D, Xiang M, Luo M, Dong Y, Fang Y, Zhang S, et al. Using gamification and social incentives to increase physical activity and related social cognition among undergraduate students in Shanghai, China. Int J Environ Res Public Health 2019;16(5):858. [doi: 10.3390/ijerph16050858]
40. Tsay CH, Kofinas AK, Trivedi SK, Yang Y. Overcoming the novelty effect in online gamified learning systems: an empirical evaluation of student engagement and performance. Comput Assist Lang Learn 2019;36(2):128-146. [doi: 10.1111/jcal.12385]
41. Sharma K, Nguyen A, Hong Y. Self ‐ regulation and shared regulation in collaborative learning in adaptive digital learning environments: a systematic review of empirical studies. Brit J Educational Tech 2024;55(4):1398-1436. [doi: 10.1111/bjet.13459]
42. Ribeiro E, Farias C, Mesquita I. “The Game Changers”: How equity-driven pedagogical scaffolding reduces participation disparities in physical education. Educ Sci 2024;14(10):1077. [doi: 10.3390/educsci14101077]
43. Rincon-Flores EG, Castano L, Guerrero Solis SL, Olmos Lopez O, Rodríguez Hernández CF, Castillo Lara LA, et al. Improving the learning-teaching process through adaptive learning strategy. Smart Learn Environ 2024;11(1):27. [doi: 10.1186/s40561-024-00314-9]
44. Boke H, Aygun Y, Tufekci S, Yagin FH, Canpolat B, Norman G, et al. Effects of cooperative learning on students' learning outcomes in physical education: a meta-analysis. Front Psychol 2025;16:1508808. [doi: 10.3389/fpsyg.2025.1508808] [Medline: 40432799]
45. Tariq MU, Sergio RP. Innovative assessment techniques in physical education: exploring technology-enhanced and student-centered models for holistic student development. In: Global Innovations in Physical Education and Health. Hershey, PA: IGI Global; 2025:85-112.
46. Fernandez-Rio J, Iglesias D. What do we know about pedagogical models in physical education so far? An umbrella review. Phys Educ Sport Pedagogy 2022;29(2):190-205. [doi: 10.1080/17408989.2022.2039615]
47. Sun Z, Yuan Y, Xiong X, Meng S, Shi Y, Chen A. Predicting academic achievement from the collaborative influences of executive function, physical fitness, and demographic factors among primary school students in China: ensemble learning methods. BMC Public Health 2024;24(1):274 [FREE Full text] [doi: 10.1186/s12889-024-17769-7] [Medline: 38263081]
48. Wang X, Yang L, Chen K, Zheng Y. Understanding teacher emotional exhaustion: exploring the role of teaching motivation, perceived autonomy, and teacher-student relationships. Front Psychol 2023;14:1342598 [FREE Full text] [doi: 10.3389/fpsyg.2023.1342598] [Medline: 38259554]
49. Faber TJE, Dankbaar MEW, van den Broek WW, Bruinink LJ, Hogeveen M, van Merriënboer JJG. Effects of adaptive scaffolding on performance, cognitive load and engagement in game-based learning: a randomized controlled trial. BMC Med Educ 2024;24(1):943 [FREE Full text] [doi: 10.1186/s12909-024-05698-3] [Medline: 39210381]
50. Xu L, Zou X, Hou Y. Effects of feedback visualisation of peer ‐ assessment on pre ‐ service teachers' data literacy, learning motivation, and cognitive load. J Comput Assist Learn 2024;40(4):1447-1462. [doi: 10.1111/jcal.12955]

51. Carcelén-Fraile MDC. Active gamification in the emotional well-being and social skills of primary education students. Educ Sci 2025;15(2):212. [doi: 10.3390/educsci15020212]
52. Hsia LH, Hwang GJ, Hwang JP. AI-facilitated reflective practice in physical education: an auto-assessment and feedback approach. Interact Learn Environ 2023;32(9):5267-5286. [doi: 10.1080/10494820.2023.2212712]
53. Kam AHT, Umar IN. Fostering autonomous motivation: a deeper evaluation of gamified learning. J Comput High Educ 2023;36(2):368-388. [doi: 10.1007/s12528-023-09358-1]
54. De Meester A, Galle J, Soenens B, Haerens L. Perseverance in motor tasks: the impact of different types of positive feedback. Phys Educ Sport Pedagogy 2022;29(2):221-234. [doi: 10.1080/17408989.2022.2054969]
55. Castle N, Howells K. A model for effective learning in competition: a pedagogical tool to enhance enjoyment and perceptions of competency in physical education lessons for young children. Children (Basel) 2024;11(1):111 [FREE Full text] [doi: 10.3390/children11010111] [Medline: 38255424]
56. Gan Y, Yang Y, Wu Y, Li H, Zhang Y. The effects of teacher persuasion on Chinese junior high school students' math self-efficacy: the multiple mediation of intrinsic value and math anxiety. Curr Psychol 2025;44(5):3956-3968. [doi: 10.1007/s12144-025-07381-2]
57. Sal-de-Rellán A, Hernández-Suárez Á, Hernaiz-Sánchez A. Gamification and motivation in adolescents. Systematic review from physical education. Front Psychol 2025;16:1575104 [FREE Full text] [doi: 10.3389/fpsyg.2025.1575104] [Medline: 40196203]
58. Spaho E, Çiço B, Shabani I. IoT integration approaches into personalized online learning: systematic review. Computers 2025;14(2):63. [doi: 10.3390/computers14020063]
59. Strielkowski W, Grebennikova V, Lisovskiy A, Rakhimova G, Vasileva T. AI-driven adaptive learning for sustainable educational transformation. Sustain Dev 2024;33(2):1921-1947. [doi: 10.1002/sd.3221]
60. Saini M, Sengupta E, Sharma N. AI-assisted hand gestures based smart feedback system for educators. Educ Inf Technol 2025;30(10):13281-13308 [FREE Full text] [doi: 10.1007/s10639-025-13358-2]
61. Lin H, Chen Q. Artificial intelligence (AI) -integrated educational applications and college students' creativity and academic emotions: students and teachers' perceptions and attitudes. BMC Psychol 2024;12(1):487 [FREE Full text] [doi: 10.1186/s40359-024-01979-0] [Medline: 39285268]
62. Fang W, Fan S, Zheng H, Fang Z, You Y, Yin B, et al. How to improve mental health in the older adults through AI-enhanced physical activity: an emerging research topic. Humanit Soc Sci Commun 2025;12(1):1-15. [doi: 10.20944/preprints202502.2037.v1]
63. Singh B, Bennett H, Miatke A, Dumuid D, Curtis R, Ferguson T, et al. Effectiveness of exercise for improving cognition, memory and executive function: a systematic umbrella review and meta-meta-analysis. Br J Sports Med 2025;59(12):866-876 [FREE Full text] [doi: 10.1136/bjsports-2024-108589] [Medline: 40049759]
64. Furley P, Schütz LM, Wood G. A critical review of research on executive functions in sport and exercise. Int Rev Sport Exerc Psychol 2023;18(1):316-344. [doi: 10.1080/1750984x.2023.2217437]
65. Gkintoni E, Vantaraki F, Skoulidi C, Anastassopoulos P, Vantarakis A. Gamified health promotion in schools: the integration of neuropsychological aspects and CBT-a systematic review. Medicina (Kaunas) 2024;60(12):2085 [FREE Full text] [doi: 10.3390/medicina60122085] [Medline: 39768964]
66. Nguyen-Viet B, Nguyen-Duy C, Nguyen-Viet B. How does gamification affect learning effectiveness? The mediating roles of engagement, satisfaction, and intrinsic motivation. Interact Learn Environ 2024;33(3):2635-2653. [doi: 10.1080/10494820.2024.2414356]
67. Mercan G, Selçuk ZV. Investigating the impact of game-based learning and gamification strategies in physical education: a comprehensive systematic review. JIETP 2024;6(1):1-14. [doi: 10.47157/jietp.1389843]
68. Sañudo B, Sanchez-Trigo H, Domínguez R, Flores-Aguilar G, Sánchez-Oliver A, Moral JE, et al. A randomized controlled mHealth trial that evaluates social comparison-oriented gamification to improve physical activity, sleep quantity, and quality of life in young adults. Psychol Sport Exerc 2024;72:102590. [doi: 10.1016/j.psychsport.2024.102590] [Medline: 38218327]
69. Zha H, Li W, Wang W, Xiao J. The paradox of AI empowerment in primary school physical education: why technology may hinder, not help, teaching efficiency. Behav Sci (Basel) 2025;15(2):240 [FREE Full text] [doi: 10.3390/bs15020240] [Medline: 40001871]
70. Iglesias D, Fernandez-Rio J. A model fidelity check in cooperative learning research in physical education. Phys Educ Sport Pedagogy 2024:1-16 [FREE Full text] [doi: 10.1080/17408989.2024.2342831]
71. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. Berlin, Germany: Springer Science & Business Media; 2013.
72. Kassenkhan AM, Moldagulova AN, Serbin VV. Gamification and artificial intelligence in education: a review of innovative approaches to fostering critical thinking. IEEE Access 2025;13:98699-98728. [doi: 10.1109/access.2025.3576147]
73. Chen CH, Chang CL. Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior. Educ Inf Technol 2024;29(14):18621-18642. [doi: 10.1007/s10639-024-12553-x]
74. Taggart J, Wheeler LB. Collaborative learning as constructivist practice: an exploratory qualitative descriptive study of faculty approaches to student group work. Act Learn High Educ 2023;26(1):59-76. [doi: 10.1177/14697874231193938]

75. Schrank FA, Wendling, BJ. The woodcock-johnson III normative update: tests of cognitive abilities and tests of achievement. In: Flanagan DP, Harrison PL, editors. Contemporary Intellectual Assessment: Theories, Tests, and Issues. New York, NY: The Guilford Press; 2012:297-335.
76. Hamari J, Koivisto J, Sarsa H. Does gamification work? -- A literature review of empirical studies on gamification. 2014 Presented at: 47th Hawaii International Conference on System Sciences; January 6-9, 2014; Waikoloa, HI p. 3025-3034. [doi: 10.1109/hicss.2014.377]
77. Kendzierski D, DeCarlo KJ. Physical activity enjoyment scale: two validation studies. J Sport Exerc Psychol 1991;13(1):50-64. [doi: 10.1123/jsep.13.1.50]
78. Laal M, Ghodsi SM. Benefits of collaborative learning. Procedia Soc Behav Sci 2012;31:486-490. [doi: 10.1016/j.sbspro.2011.12.091]
79. Belmont M, Skinner EA, Wellborn J, Connell J. Teacher as Social Context Questionnaire. APA PsycNet. 1988. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft10488-000 [accessed 2026-01-24]
80. McAuley E, Duncan T, Tammen VV. Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: a confirmatory factor analysis. Res Q Exerc Sport 1989;60(1):48-58. [doi: 10.1080/02701367.1989.10607413] [Medline: 2489825]
81. Ifenthaler D, Eseryel D, Ge X. Assessment in Game-Based Learning. New York, NY: Springer; 2012:1-8.
82. Caspari-Sadeghi S. Artificial intelligence in technology-enhanced assessment: a survey of machine learning. J Educ Technol Syst 2022;51(3):372-386. [doi: 10.1177/00472395221138791]
83. Publication Manual of the American Psychological Association, Seventh Edition. Washington, DC: American Psychological Association; 2020.
84. Criollo-C S, Guerrero-Arias A, Guaña-Moya J, Samala A, Luján-Mora S. Towards sustainable education with the use of mobile augmented reality in early childhood and primary education: a systematic mapping. Sustainability 2024;16(3):1192. [doi: 10.3390/su16031192]
85. Falola H, Joel OO, Akinbode J, Ojebola O. Emotional intelligence and academic engagement of graduate students: the moderating role of digital learning support. Afr J Econ Manag Stud 2025:1-20 [FREE Full text] [doi: 10.1108/ajems-03-2024-0203]
86. Gkintoni E, Vantaraki F, Skoulidi C, Anastassopoulos P, Vantarakis A. Promoting physical and mental health among children and adolescents via gamification-a conceptual systematic review. Behav Sci (Basel) 2024;14(2):102 [FREE Full text] [doi: 10.3390/bs14020102] [Medline: 38392455]
87. Chaiyarat K. Enhancing creative problem solving and learning motivation in social studies classrooms with gamified cooperative learning. Think Ski Creat 2024;54:101616. [doi: 10.1016/j.tsc.2024.101616]
88. Aibar-Almazán A, Castellote-Caballero Y, Carcelén-Fraile MDC, Rivas-Campo Y, González-Martín AM. Gamification in the classroom: Kahoot! As a tool for university teaching innovation. Front Psychol 2024;15:1370084 [FREE Full text] [doi: 10.3389/fpsyg.2024.1370084] [Medline: 38646120]
89. Trajkovski G, Hayes H. AI-assisted formative assessment and feedback. In: AI-Assisted Assessment in Education. Cham: Palgrave Macmillan; 2025:283-312.
90. To J. Feedback partnerships: strengthening students' proactive recipience through co-creating dialogic feedback. High Educ Res Dev 2025;44(4) [FREE Full text] [doi: 10.1080/07294360.2024.2445570]
91. Cha W, Hong M, Glassman M, Anderman EM, Lin T. The roles of technology efficacy and networking agency in elementary students' engagement in online and face‐to‐face technology‐mediated learning. Brit J Educational Tech 2025;56(6):2623-2646 [FREE Full text] [doi: 10.1111/bjet.13597]
92. Ryan RM, Deci EL. Self-determination theory. In: Maggino F, editor. Encyclopedia of Quality of Life and Well-Being Research. Cham: Springer; 2023.
93. Chen C. Exploring the impact of acute physical activity on creative thinking: a comprehensive narrative review with a focus on activity type and intensity. Discov Psychol 2024;4(1):3. [doi: 10.1007/s44202-024-00114-9]
94. Chen BC. Using Vygotsky's sociocultural theory to explore ethnic cultural representation in Taiwanese preschool children's play. Front Educ 2025;10:1569322. [doi: 10.3389/feduc.2025.1569322]
95. Qi S, Derakhshan A. Technology-based collaborative learning: EFL learners' social regulation and modifications in their academic emotions and academic performance. Educ Inf Technol 2024;30(7):8611-8636. [doi: 10.1007/s10639-024-13167-z]
96. Banik BG, Gullapelly A. AI-powered gamification and interactive learning tools for enhancing student engagement. In: InDriving Quality Education Through AI and Data Science. Hershey, PA: IGI Global Scientific Publishing; 2025:283-310.
97. Dinh CT, Phuong HY. MOOC learners' perspectives of the effects of self-regulated learning strategy intervention on their self-regulation and speaking performance. Cogent Education 2024;11(1):2378497. [doi: 10.1080/2331186x.2024.2378497]
98. Suresh Babu S, Dhakshina Moorthy A. Application of artificial intelligence in adaptation of gamification in education: a literature review. Comp Applic In Engineering 2023;32(1):e22683. [doi: 10.1002/cae.22683]
99. Ng DTK, Xinyu C, Leung JKL, Chu SKW. Fostering students' AI literacy development through educational games: AI knowledge, affective and cognitive engagement. Computer Assisted Learning 2024;40(5):2049-2064. [doi: 10.1111/jcal.13009]

100. Shalgimbekova K, Smagliy T, Kalimzhanova R, Suleimenova Z. Innovative teaching technologies in higher education: efficiency and student motivation. Cogent Education 2024;11(1):2425205. [doi: 10.1080/2331186x.2024.2425205]
101. Shekh-Abed A. Metacognitive self-knowledge and cognitive skills in project-based learning of high school electronics students. Eur J Eng Educ 2024;50(1):214-229. [doi: 10.1080/03043797.2024.2374479]
102. Barz N, Benick M, Dörrenbächer-Ulrich L, Perels F. The effect of digital game-based learning interventions on cognitive, metacognitive, and affective-motivational learning outcomes in school: a meta-analysis. Rev Educ Res 2023;94(2):193-227. [doi: 10.3102/00346543231167795]
103. Hsia LH, Lin YN, Lin CH, Hwang GJ. Effectiveness of gamified intelligent tutoring in physical education through the lens of self-determination theory. Comput Educ 2025;227:105212. [doi: 10.1016/j.compedu.2024.105212]
104. Bhardwaj V, Zhang S, Tan YQ, Pandey V. Redefining learning: student-centered strategies for academic and personal growth. Front. Educ 2025;10:1518602. [doi: 10.3389/feduc.2025.1518602]
105. Plass JL, Froehlich F. The future of personalized learning with AI. Learn Individ Differ 2025;124:102813. [doi: 10.1016/j.lindif.2025.102813]
106. Huseinović L. The effects of gamification on student motivation and achievement in learning english as a foreign language in higher education. MAPEH 2023;4(1):10-36. [doi: 10.53880/2744-2373.2023.4.10]
107. Kalogiannakis M, Papadakis S, Zourmpakis A. Gamification in science education. A systematic review of the literature. Educ Sci 2021;11(1):22. [doi: 10.3390/educsci11010022]
108. Ferriz-Valero A, Østerlie O, García Martínez S, García-Jaén M. Gamification in physical education: evaluation of impact on motivation and academic performance within higher education. Int J Environ Res Public Health 2020;17(12):4465 [FREE Full text] [doi: 10.3390/ijerph17124465] [Medline: 32575919]
109. Sotos-Martínez VJ, Ferriz-Valero A, García-Martínez S, Tortosa-Martínez J. The effects of gamification on the motivation and basic psychological needs of secondary school physical education students. Phys Educ Sport Pedagogy 2022;29(2):160-176. [doi: 10.1080/17408989.2022.2039611]
110. Biswas U, Bhattacharya S. ML-based intelligent real-time feedback system for blended classroom. Educ Inf Technol 2023;29(4):3923-3951. [doi: 10.1007/s10639-023-11949-5]
111. Qadir HM, Khan RA, Rasool M, Sohaib M, Shah MA, Hasan MJ. An adaptive feedback system for the improvement of learners. Sci Rep 2025;15(1):17242 [FREE Full text] [doi: 10.1038/s41598-025-01429-w] [Medline: 40383780]

Abbreviations

AI: artificial intelligence
AVE: average variance extracted
CFA: confirmatory factor analysis
CFI: comparative fit index
CLT: cognitive load theory
CR: composite reliability
HTMT: heterotrait-monotrait
PE: physical education
RMSEA: root-mean-square error of approximation
SDT: self-determination theory
SEM: structural equation modeling
TLI: Tucker-Lewis Index

Edited by S Brini; submitted 22.Jul.2025; peer-reviewed by YC Loh, FMJM Shamrat; comments to author 26.Oct.2025; accepted 06.Jan.2026; published 02.Feb.2026.

Please cite as:
Zhang J, Oh SS, Xu Y
Gamified Physical Education and Cognitive Performance Among Chinese Secondary School Students: Cross-Sectional Moderation Mediation Study
JMIR Serious Games 2026;14:e81086
URL: https://games.jmir.org/2026/1/e81086
doi:10.2196/81086
PMID:

Original Paper

Effects of Exergaming Tennis on Players' Tennis Skills and Mental State Compared to Regular Tennis in Adult Players: Quasi-Experimental Study

Jia Sheng Ngo[1], MMedSc; See Ziau Hoe[1], PhD; Maziah Mat Rosly[1], PhD

Department of Physiology, Faculty of Medicine, Universiti Malaya, Kuala Lumpur, Malaysia

Corresponding Author:
Maziah Mat Rosly, PhD
Department of Physiology
Faculty of Medicine
Universiti Malaya
Kuala Lumpur, 50603
Malaysia
Phone: 60 379674924
Email: maziahmr@um.edu.my

Abstract

Background: Exergaming, which combines physical activity with interactive gaming, has been shown to improve motor skills and fitness. However, exergaming's potential in complex, open-skill sports such as tennis, which require real-time coordination, decision-making, and technical precision, remains underexplored. Furthermore, only a few studies have evaluated the impact of exergaming on both technical skill development and psychological outcomes such as motivation and confidence, especially among novice players. This study addresses these gaps by comparing the combination of exergame-based tennis training and on-court tennis training (EBTT+OCTT) with on-court tennis training alone (OCTT×2) in improving technical skills, grip strength, confidence, and motivation.

Objective: This study aims to assess the effect of exergaming tennis in developing tennis technical skills, grip strength, confidence level, and motivation level.

Methods: In a 12-week quasi-experimental trial, 66 novices were randomized to either the EBTT+OCTT or the OCTT×2 group. Participants consisted of 22 males and 44 females, with the mean BMI and age of approximately 22 (SD 2.9) kg/m^2 and 26 (SD 7) years, respectively. The EBTT+OCTT group had 1 weekly session of OCTT and 1 session of exergaming training using Virtual Tennis (PlayStation 3 Move), while the OCTT×2 group completed 2 weekly on-court sessions. Outcomes included tennis technical skills such as groundstroke technique, service speed, and service accuracy (assessed using the Hewitt tennis test), grip strength (using a handheld dynamometer), confidence (using the Sports Confidence Inventory), and motivation (using the Sport Motivation Scale). Mean differences (MDs) between pre- and postintervention were collected and analyzed using repeated-measures ANOVA (α=.05) and Pearson correlation analysis. Effect sizes were calculated using partial eta squared (ηp^2), where values ≥0.14 indicated large effects.

Results: After 12 weeks of training, both EBTT+OCTT and OCTT×2 groups showed significant improvements (P<.001) in tennis technical skills (MD=17.06-22.62), grip strength (MD=9.59-11.04 kg), and confidence levels (MD=23.29-26.28). These outcome measures have large effect sizes (ηp^2=0.84-0.92); however, they did not significantly differ (P>.05) across the groups when compared, with $P_{grip\ strength}$=.24, P_{hit}=.97, $P_{accuracy}$=.86, $P_{speed=}$.72, and $P_{confidence}$=.31. In terms of motivation, EBTT+OCTT retained intrinsic motivation (IM) better than OCTT×2, with significant reductions in IM, mainly IM-to-know (MD=7, SD 2.95) and IM-to-accomplish (MD=5, SD 3.77) observed in the OCTT×2 group (P<.001). Grip strength, confidence, and motivation levels (except amotivation) showed positive correlations with tennis technical skills (r=0.39-0.80).

Conclusions: EBTT+OCTT and OCTT×2 significantly improve tennis skills and confidence levels in novice players, although no significant differences were found between the two. However, EBTT+OCTT appeared to better sustain IM. Thus, EBTT+OCTT may serve as a supplementary tool for novice players to better learn tennis.

(JMIR Serious Games 2026;14:e73732) doi:10.2196/73732

KEYWORDS

exergaming; tennis; sports skills; psychology; sport training

Introduction

Technological innovations have facilitated the integration of physical activity through electronic platforms, particularly via exergaming—interactive video games that require physical movement for gameplay [1]. Substantial evidence demonstrates the efficacy of exergaming in enhancing motor coordination, reaction time, and physical fitness across various populations, including pediatric, geriatric, rehabilitation, and special needs cohorts [2,3]. These systems present novel opportunities for incorporating exercise into daily life, prompting investigation of their applications in sports training. According to the existing literature, most research has focused primarily on endurance and closed-skill sports (eg, cycling, running, and rowing), as their predictable, self-paced nature translates effectively to virtual environments [4-6]. In contrast, open-skill sports such as tennis have received limited attention due to the technical challenges of accurately replicating their complex biomechanics and decision-making demands in digital platforms [4]. Tennis demands precise mind-body coordination, decision-making skills, and real-time reactions [7]. While exergaming shows potential for early skills acquisition, the current literature lacks studies examining its effects in skill-based sports.

Tennis is a complex, open-skill sport that demands not only physical skill and execution of techniques such as serving and groundstrokes but also strong psychological attributes, including confidence and motivation [7,8]. Technical proficiency, such as accurate and powerful strokes, is fundamental to performance, yet these motor skills do not operate in isolation, and physical capability, such as grip strength, may also play a role in ensuring these skills are executed appropriately [4,9,10]. In addition, psychological readiness can influence how consistently and effectively players perform under pressure, make decisions, and maintain effort during training or competition [7,11].

While physical outcomes have been moderately investigated, there is a lack of research examining psychological outcomes, particularly confidence and motivation, which are known to influence motor learning and sports performance [4,7,11]. Confidence, for example, plays a key role in motor learning and performance; players with greater sport confidence are more likely to take risks, recover from mistakes, and execute techniques successfully under stress [7,12]. Similarly, motivation, especially intrinsic motivation (IM), affects the degree of engagement, persistence, and enjoyment during practice, which can directly enhance motor learning and skill retention [7,13]. These psychological factors are not only relevant to sport performance but may also be interrelated with technical skill development, making it important to study both domains simultaneously [7,11-13]. For instance, a motivated and confident player may be more willing to engage in deliberate practice, which leads to better technical improvement. Only a limited number of studies have concurrently examined the interrelationship between technical and psychological skills in an exergaming context, leaving a gap in understanding how these domains interact to support performance and engagement.

Exergaming, by blending physical movement with gaming elements, has the potential to impact both domains [1]. It can foster motor skill repetition in a stimulating environment while simultaneously supporting motivation and self-confidence through goal-setting and feedback [3]. However, most previous exergaming research has focused either on general fitness or cognitive outcomes, with limited exploration of its impact on sport-specific technical skills and psychological outcomes in a combined manner.

While both exergaming and traditional on-field exercises aim to promote physical fitness, they differ considerably in context, delivery, and user engagement. Research has consistently demonstrated that on-field exercises such as tennis, football, and badminton contribute to improved cardiovascular fitness, enhanced muscular strength, better body composition, and overall physical and mental well-being [8,14]. However, despite their benefits, on-field exercises can present certain challenges, such as time constraints, limited access to appropriate facilities, weather conditions, and personal motivation [14-16]. Exergaming, in contrast, combines physical movement with interactive digital gameplay, making exercise more accessible, enjoyable, and gamified [3]. Research has shown that adding exergaming to exercise regimes such as badminton, tennis, and golf can help encourage social interaction, allowing users to participate in cooperative or competitive activities with family, friends, or online communities, which may improve the appeal and sustainability of physical activity [17,18].

Tennis was chosen in this study because there is limited research examining this sport. There was also a need to evaluate whether skills learned through exergaming can transfer to real tennis performance. Although understanding the rules and strategies of tennis is important, mastery of the sport requires repeated physical practice to develop technical proficiency, coordination, and real-time decision-making skills. Exergaming may help players develop motor skills to respond to service and shots through repetitive practice, combining mental, visual, and motor learning.

Thus, the main objectives of this research were to determine whether integrating exergame-based tennis training and on-court tennis training (EBTT+OCTT) can help improve tennis technical skills, grip strength, confidence, and motivation levels compared with only the traditional on-court tennis training (OCTT×2). In addition, this research aimed to determine whether there is any correlation between grip strength, confidence, and motivation, to better understand their interrelationship and impact on tennis performance.

Methods

Study and Trial Designs

This quasi-experimental study was conducted in Malaysia, where participants were initially randomly allocated to either the EBTT+OCTT group or the OCTT×2 group as a parallel-group experimental study (aged 18 to 40 years, with balanced randomization in a ratio of 1:1). However, about 9%-10% (OCTT×2 [n=3] and EBTT+OCTT [n=4]) of participants in each group requested to change their assigned allocation, and these requests were accommodated to minimize participant dropout. These requests were made before commencement of the experimental procedure, and no participants were switched during the experimental period.

Participants

Social media postings and word-of-mouth marketing were used to recruit research participants. Participants were predominantly urban residents, consisting of university students and young professionals from diverse socioeconomic strata. The sample size calculator program G*Power (version 3.1.9.4; Heinrich Heine University) was used to determine the sample size. A repeated measures ANOVA (within- and between-groups' interactions) was selected as the statistical test. The calculation was based on a large effect size (f=0.4), a significance level of 0.05, and a power (1−β1) of 0.95 [19]. The large effect size was chosen based on previous literature in sports and exergaming interventions reporting medium to large effect sizes for motor and psychological outcomes [20,21]. In addition, as this study was exploratory and involved multiple dependent variables with expected training-related gains, a conservative large effect size was used to ensure adequate statistical power. Two groups (EBTT+OCTT and OCTT×2) and 2 measurement time points (pre- and posttraining) were included, with an estimated correlation of 0.5 between repeated measures. The analysis determined that a minimum of 24 participants (12 per group) were required. To account for a potential dropout rate of 20%, 29 participants were required in total. This study recruited 66 participants in total, with 33 participants assigned to each group (Textbox 1).

Textbox 1. The inclusion and exclusion criteria of the study.

Inclusion criteria

- Individuals with less than 1 year of tennis experience.
- Aged between 18 and 40 years.
- No medical history or surgical history.
- Physical activity level below moderate as assessed by the International Physical Activity Questionnaire (IPAQ).

Exclusion criteria

- Individuals who are obese (BMI>30 kg/m^2).
- Individuals who do not understand English or Malay.

In the recruitment process, 74 individuals were screened, and 66 met the inclusion criteria. The others were excluded for reasons such as having over a year of tennis experience (3 individuals), a high International Physical Activity Questionnaire (IPAQ) score (2 individuals), a BMI exceeding 30 kg/m² (2 individuals), or cardiorespiratory conditions such as asthma (1 individual). The final sample (22 males and 44 females with a mean age of 26.3 years) represented the university-affiliated and local recreational tennis community, which is roughly 60% female and predominantly aged 18 to 35 years. The participants' weekly physical activity levels, averaging about 760 metabolic equivalents of task minutes on the IPAQ scale, aligned with those of a low-to-moderately active urban adult population. The low to moderate IPAQ score was selected because it can enhance the responsiveness to training interventions, as individuals with lower activity levels are more likely to demonstrate measurable improvements in physical and psychological outcomes [4,7]. Furthermore, exergaming interventions have been primarily designed for and shown to be particularly beneficial in low-active or novice populations, supporting the appropriateness of this inclusion criterion [7,22]. A standardized IPAQ score was also required to reduce variability in fitness and motor performance that might influence the responsiveness to training, particularly since the study targeted novice players [22].

Experimental Procedure

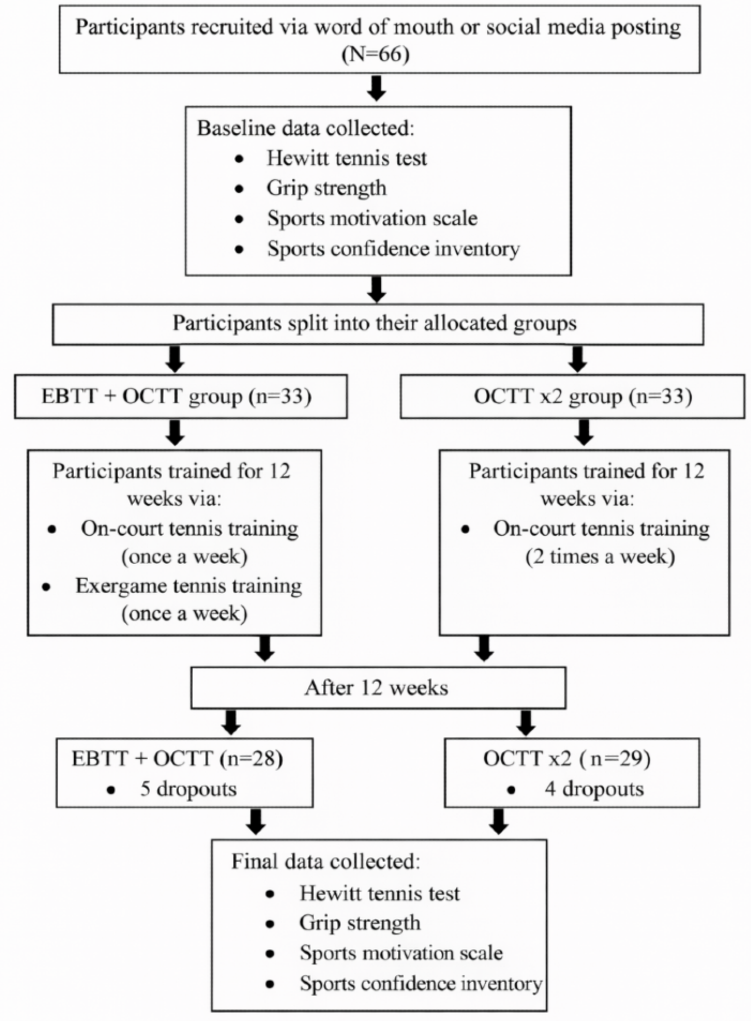
Sixty-six individuals were divided into two groups, with a total of 33 individuals in each of the OCTT x 2 and EBTT+OCTT groups. The participants in the EBTT+OCTT group practiced tennis on both the exergaming platform and the traditional tennis court, whereas the OCTT×2 group practiced only on the traditional tennis court. Before the training started, the participants' consent, demographic data, physical activity level via IPAQ, and baseline data such as sports motivation test, sports confidence inventory, grip strength via handheld dynamometer, and technical skills via the Hewitt tennis test were collected. All questionnaires and outcome measures were valid and reliable [22-26].

Both groups followed identical protocols for their on-court training sessions, which were conducted on tennis courts and with equipment meeting International Tennis Federation regulatory standards [27]. The sessions were done for 1.5 hours, comprising a warm-up period of 15 minutes, a 1-hour main practice, and a 15-minute cool-down. The exercise session duration and frequency were set following the American College of Sports Medicine guidelines [28,29]. The OCTT×2 group practiced on-court tennis twice a week, whereas the EBTT+OCTT group had one on-court and one exergaming session per week. In addition to the on-court tennis practice, participants in the EBTT+OCTT group also practiced tennis using a video game version of Virtual Tennis (Sports Champion

2, Zindagi Games, San Diego Studio) on a gaming device, the PlayStation 3 Move, for 1.5 hours.

Training included swing practices such as service, forehand, and backhand, and matches (either with a virtual avatar or among themselves). Matches were conducted between the 10th and 12th weeks. Both groups performed the task for 12 weeks (24 practice sessions). There was no group switching among participants during the experimental period. Four individuals in the OCTT×2 group and 5 in the EBTT+OCTT group dropped out during the second and third weeks, respectively, during the 12-week training period. Figure 1 provides the flowchart of the experimental design.

Figure 1. Flowchart of the quasi-experimental study. EBTT: exergame-based tennis training; OCTT: on-court tennis training.



Outcome Measures

Measurement of Tennis Technical Skills

In this study, the Hewitt tennis test was used to assess the player's technical skills, such as service speed, service accuracy, and tennis strokes (forehand and backhand). This test was done pre- and post-12 weeks of practice. It has been demonstrated that this test is reliable and valid to conduct [25]. Participants in the tennis hit test aimed to score points by hitting balls into specific zones over the net, using forehand or backhand strokes during 10 trials. In the service accuracy and speed test, they served 10 balls into a target area, with accuracy scores based on placement and speed determined by the ball's bounce distance after hitting the service court. Each subtest score is:

- Groundstroke (hit) score: participants earned 1-5 points per stroke based on where the ball landed on a court target zone. The maximum score for this test was 50.
- Service accuracy score: participants were instructed to serve into marked target zones on the court. Points were awarded based on the accuracy of placement, with a maximum of 50 points.
- Service speed score: speed was estimated based on the bounce distance of the ball in the service court, categorized into point zones, with a maximum possible score of 50.

All outcomes were treated as continuous variables (range: 0-50) and used in the statistical analysis. The normal value given by Hewitt's Tennis test scores (score of 1-50) for beginners was graded from A (>18 points) to F (<3 points), where "A" represents an excellent score and "F" represents a very poor score [25]. While scoring the participant, the assessor was aware of the participant's group allocation.

Measurement of Grip Strength

Measuring grip strength with handheld dynamometers was used due to their high validity and reliability [30,31]. Standardized protocols were followed to ensure accurate and comparable data. This typically involves a seated position with elbows flexed at 90 degrees, neutral wrists, and the handgrip aligned with the forearm [30,31]. Following the standardized protocol, 3 attempts were made for the dominant hand with rest intervals in between attempts. The highest value in kilograms (kg) was recorded and used for analysis. Grip strength was treated as a continuous variable in the statistical analysis. Measurements were taken pre- and post-12 weeks of training. The normal value for grip strength in Malaysia for males aged 18 to 54 ranged from 37 kg to 42 kg, whereas for females of a similar age, it ranged from 22 kg to 25 kg [32,33].

Measurement of Motivation and Confidence Level

The Sports Motivation Test (SMT) and the Sports Confidence Inventory (SCI) were used to measure motivation and confidence levels of participants when playing sports pre- and post-12 weeks of training. The SMT, developed by Pelletier et al [34], is a self-report questionnaire that assessed 7 types or subscales of motivation: IM-to-know, IM-to-accomplish, IM-to-experience stimulation, extrinsic motivation (EM)–identified, EM-introjected, EM–external regulation, and amotivation. Participants indicated the extent to which each statement reflected their motivation for engaging in sports, ranging from "not at all true" to "very true." Each subscale score was calculated as the sum of all 4 items belonging to that subscale, resulting in 7 separate subscale scores ranging from 4 to 28. These subscale scores were used as separate continuous outcome variables in the analysis of motivation change over time and in correlation with tennis performance. The SMT is widely used in research settings to examine athletes' motivation and understand the factors that drive their engagement in sports. High internal consistency is indicated by Cronbach alpha coefficients ranging from 0.70 to 0.90 [24,34].

The SCI, developed by Vealey [35], is a widely used questionnaire designed to assess athletes' self-confidence in sports situations. Participants rate their level of confidence in each statement on a scale ranging from "strongly disagree" to "strongly agree." The SCI measures both general and sport-specific self-confidence across various domains, such as skill execution, physical conditioning, and resiliency [23]. The total confidence score ranges from 8 to 72, with higher scores indicating greater sport-specific confidence. All 8 items were summed to calculate a total confidence score, which was used in pre-post comparisons and correlation analyses as a continuous variable. According to reliability assessments, the SCI has a high degree of internal consistency, with Cronbach alpha values between 0.70 and 0.90 [35]. Test-retest reliability has also been established, demonstrating stability over time.

Data Analysis

SPSS software, created by International Business Machines, was used to gather and analyze the data. Normality was evaluated using the Shapiro-Wilk test, which is highly effective at detecting nonnormality [36]. For normally distributed data, repeated measures ANOVA was performed. Final analysis was conducted per protocol outlined in the "Data Analysis" section, where participants were analyzed according to the group they actually participated in, instead of their original randomized groups. In addition, a modified intention-to-treat (ITT) analysis was also performed, in which all participants who were randomized initially and completed the study were included in the analysis according to their originally assigned groups. This analytical approach excludes missing data from participants who dropped out of their respective interventions. To evaluate whether the treatment effects differed significantly between modified ITT analyses and per-protocol analyses, treatment effects from both approaches were compared. The primary comparison was based on the magnitude and 95% CIs of the treatment effects. Consistency between modified ITT and per-protocol analyses effects was interpreted as evidence of robustness. Effect sizes were evaluated using partial eta squared (ηp^2), with results classified as small (0.01-0.06), medium (0.06-0.14), or large (≥0.14) [37]. Partial eta squared (ηp^2) was used as the measure of effect size for the repeated-measures ANOVA results, as it is appropriate for estimating effect sizes in factorial and within-subject designs. Unlike Cohen *d*, which is suited for simple 2-group comparisons, ηp^2 provides a proportion of variance explained by each effect (main and interaction), making it more suitable for analyzing group × time interactions in repeated-measures designs [37]. If the time (pre-post training) and group interaction are significant, posthoc pairwise comparisons were used; however, if only the time effect is significant, but there is no interaction between time and group, paired *t* tests were used to confirm improvement or mean difference (MD) in each group. A 95% CI of the MD was also calculated to determine the range within which the true difference between population means is likely to fall. The correlation between the changes (differences between pre- and postintervention) in players' grip strength, confidence, and motivation levels with their tennis technical skills was analyzed using Pearson's product-moment correlation coefficient (*r*). According to Cohen [38], the strength of the relationship is weak (*r*=0.10), moderate (*r*=0.30), or strong (r=0.50). Pearson *r* value was squared and multiplied by 100 (r^2) to determine the percentage of variation explained by 1 variable for another. In this study, r^2 was used to determine the percentage of variability in tennis technical skills accounted for by grip strength, confidence, and motivation levels. Findings with a $P<.05$ were considered statistically significant.

Ethical Considerations

This study involving human participants was reviewed and approved by the Universiti Malaya Research Ethics Committee, Malaysia. The ethics approval was granted on January 26, 2023 (UM.TNC2/UMREC_2286). Written informed consent was obtained from all participants before enrollment in the study. Participants were informed about the study objectives, procedures, potential risks and benefits, and their right to withdraw at any time without penalty. The study did not involve any secondary use of existing data; all data were collected prospectively as part of this trial. To ensure participant privacy and data confidentiality, all data collected remained anonymous by assigning unique participant codes. Personal identifiers were not linked to the study data. Electronic records were stored securely on password-protected devices. Participants did not receive any monetary compensation for their participation. However, they were provided with gift vouchers from local grocery stores and restaurants and were allowed to retain access to the tennis training facilities during the study period at no cost, as a token of appreciation for their involvement. Written informed consents were also obtained from participants when their images were used in the papers. This trial was not preregistered, as it did not measure any clinical outcomes.

Results

Overview

The results are provided in 2 parts: the first part is the comparative analysis of the physical variables (tennis technical skills and grip strength) and psychological variables (confidence and motivation levels) between EBTT+OCTT and OCTT×2, and the second part is the correlation analysis of the relationships between grip strength, confidence, and motivation with the tennis technical skills. All variables are presented as mean and SD for parametric data. Demographic data, including BMI, age, sex, and physical activity level, were collected. Independent *t* tests revealed no significant differences between groups in age, training experience, sex (male and female), BMI, and physical activity levels (Table 1).

Table 1. Summary of demographic data for the quasi-experimental study.

Parameter	EBTT+OCTT[a] (n=33)	OCTT×2[b] (n=33)	*P* value
Age (y), mean (SD)	26.82 (7.08)	25.85 (6.15)	.64
Sex, n (%)			
Male	10 (30)	12 (36)	.68
Female	23 (70)	21 (64)	.79
Training experience (mo), mean (SD)	3.3 (2.44)	2.9 (3.12)	.86
Height (m), mean (SD)	1.64 (0.66)	1.65 (0.58)	.34
Weight (kg), mean (SD)	60.1 (7.8)	59.6 (7.7)	.54
BMI (kg/m^2), mean (SD)	22.35 (2.94)	21.89 (2.90)	.57
International Physical Activity Questionnaire (METs[c]), mean (SD)	808.94 (300.97)	712.35 (311.81)	.43
Dropouts, n (%)	5 (15)	4 (12)	—[d]

[a]EBTT+OCTT: exergame-based tennis training with on-court tennis training.

[b]OCTT: on-court tennis training.

[c]MET: metabolic equivalent of task.

[d]Not applicable.

Data collections were conducted from January to April 2024. During the 2nd and 3rd weeks of the study, 4 participants from the OCTT×2 group and 5 from the EBTT+OCTT group dropped out due to work schedules, health issues, transportation difficulties, or family emergencies (Table 1). The data of participants who dropped out were not used in the analysis. Shapiro-Wilk normality tests indicated that the differences between paired variables in the EBTT+OCTT and OCTT×2 groups were not statistically significant (*P*>.05), confirming that all paired variables were normally distributed.

In general, the treatment effect estimates from the modified ITT and per-protocol analyses were generally consistent for all outcome measures. The outcome measures showed overlapping 95% CIs and similar significance levels (P) between the 2 analyses, indicating no meaningful difference between the 2 analytical approaches. The table on the modified ITT analysis is provided in Multimedia Appendix 1.

Effects of EBTT+OCTT and OCTT×2 on Physical and Psychological Variables

Effect of EBTT+OCTT and OCTT×2 on Grip Strength

Both groups showed a significant improvement (*P*<.001) with a large effect size (ηp^2=0.84) from pre- to posttraining (Table 2). A total of 28 of 33 participants in the EBTT+OCTT group have an increase in grip strength by 1.5 times or 11.04 kg on average, while 29 of 33 participants in the OCTT×2 group improved their grip strength by 1.6 times or an average of 9.59 kg. However, grip strength showed no significant difference between groups (*P*=.24) with a small effect size (ηp^2=0.03).

Table 2. Comparison of grip strength, hit scores, accuracy scores, speed scores, and confidence level within and between groups.

Outcome measure and group	Pretest, mean (SD)	Posttest, mean (SD)	Value, MD[a] (95% CI; % changed)	Within-group *P* value	Between groups *P* value
Grip strength				<.001	.24
EBTT+OCTT[b]	21.68 (14.96)	32.71 (0.82)	11.04 (9.24-12.83; 51.10)		
OCTT[c]×2	16.10 (13.97)	25.69 (0.68)	9.59 (7.86-11.31; 59.57)		
Hit scores				<.001	.97
EBTT+OCTT	4.32 (4.68)	21.46 (7.21)	17.14 (14.95-19.34; 400)		
OCTT×2	5.31 (5.41)	22.38 (9.23)	17.06 (14.27-19.86; 321.28)		
Accuracy scores				<.001	.86
EBTT+OCTT	4.21 (3.25)	26.54 (8.33)	22.32 (19.55-25.09; 530.17)		
OCTT×2	4.66 (3.30)	27.28 (7.51)	22.62 (20.38-24.86; 485.41)		
Speed scores				<.001	.72
EBTT+OCTT	3.50 (2.92)	21.25 (6.03)	17.75 (15.62-19.88; 507.14)		
OCTT×2	2.97 (2.16)	20.17 (6.34)	17.21 (14.97-19.44; 579.46)		
Confidence level				<.001	.31
EBTT+OCTT	37.04 (9.21)	60.32 (2.90)	23.29 (18.93-27.64; 62.88)		
OCTT×2	34.17 (7.31)	60.45 (3.80)	26.28 (22.19-30.36; 76.91)		

[a]MD: mean difference.

[b]EBTT+OCTT: exergame-based tennis training with on-court tennis training.

[c]OCTT: on-court tennis training.

Effect of EBTT+OCTT and OCTT×2 on Technical Skills

Technical skills were categorized into 3 measures: forehand and backhand (hit) score, service accuracy, and service speed. Significant differences (P<.001) were found in all 3 measures, pre- and posttraining, for both groups, with large effect sizes: ηp^2_{hit}=0.88, $\eta p^2_{accuracy}$=0.92, and ηp^2_{speed}=0.91. In the EBTT+OCTT group, 28 of 33 participants' hit score increased by 5 times (MD 17.14), service accuracy by 6.3 times (MD 22.32), and service speed by 6.1 times (MD 17.75). In the OCTT×2 group, 29 of 33 participants' hit score increased by 4.2 times (MD 17.06), service accuracy by 5.9 times (MD 22.62), and service speed by 6.8 times (MD 17.21). However, for all technical skill outcome measures, there were no significant differences between groups (P_{hit}=.97, $P_{accuracy}$=.86, and P_{speed}=.72), with small effect sizes (ηp^2_{hit}=0.01, $\eta p^2_{accuracy}$=0.03, and ηp^2_{speed}=0.01) (Table 2).

Effect of EBTT+OCTT and OCTT×2 on Confidence Level

The main effect of time was significant (P<.001) with a large effect size (ηp^2=0.84), indicating that participants in both groups showed significant improvement from pre- to posttraining. In the EBTT+OCTT group, 28 of 33 participants improved their confidence level by 23.29 points (~1.6 times on average), whereas in the OCTT×2 group, 29 of 33 participants increased by 26.28 points (~1.8 times) compared with baseline. However, the interaction between pre-post training and the group was not significant (P=.31) with a small effect size of 0.02, suggesting no overall difference between the EBTT+OCTT and OCTT×2 groups (Table 2).

Effect of EBTT+OCTT and OCTT×2 on Motivation Level

Overall, 28 of 33 participants in the EBTT+OCTT group showed a significant decrease in IM-to-accomplish, EM-identified, and EM-introjected. Whereas 29 of 33 participants in the OCTT×2 group, other than a significant increase in amotivation level, there were significant decreases in IM-to-know, IM-to-accomplish, EM-identified, EM-introjected, and EM–external regulation. In terms of the comparison between groups, the reduction of IM-to-know score and IM-to-accomplish score was greater in the OCTT×2 group than the EBTT+OCTT group, leading to a statistical significance, P<.001, when compared. A large effect size was also found, ηp^2 (IM-to-know)=0.25, and ηp^2 (IM-to-accomplish)=0.12. The only finding in EM that was statistically significant (P<.001) with a large effect size of ηp^2=0.30 was the decrease in the EM–external regulation score of OCTT×2 in comparison to EBTT+OCTT (Table 3).

Relationship Between Grip Strength, Confidence, and Motivation With Tennis Technical Skills

Relationship Between Grip Strength and Tennis Technical Skill

The bivariate correlation between grip strength and tennis hit score was significantly positive (r=0.70; P<.001). Grip strength was also significantly positively correlated with service accuracy (r=0.45; P<.001) and service speed (r=0.42; P=.001). The proportion of variance (r^2) explained by grip strength for tennis hit score, service accuracy, and service speed was 0.50, 0.20, and 0.17, respectively.

Table 3. Comparison of different motivation levels within and between groups.

Outcome measures	Pretest, mean (SD)	Posttest, mean (SD)	Value, MD[a] (95% CI; % changed)	Within-group *P* value	Between groups *P* value
IM[b]-to-know					<.001
EBTT+OCTT[c]	24.18 (2.92)	22.39 (4.12)	–1.79 (–3.85 to 0.28; 7.40)	.04	
OCTT[d]×2	24.21 (2.53)	17.21 (2.88)	–7 (–8.43 to 5.57; 28.91)	<.001	
IM-to-accomplish					<.001
EBTT+OCTT	23.29 (2.57)	21.82 (2.21)	–1.46 (–2.90 to –0.03; 6.27)	.04	
OCTT×2	24.38 (2.21)	18.41 (2.21)	–5.97 (–7.09 to 4.85; 24.49)	<.001	
IM-to-experience stimulation					.66
EBTT+OCTT	24.50 (2.66)	24.18 (2.71)	–0.32 (–1.78 to 1.14; 1.31)	.66	
OCTT×2	24.21 (2.37)	24.31 (2.71)	0.10 (–1.26 to 1.47; 0.41)	.88	
EM[e]-identified					.42
EBTT+OCTT	24.54 (2.62)	19.11 (2.50)	–5.43 (–6.72 to –4.13; 22.13)	<.001	
OCTT×2	24.14 (2.84)	19.41 (2.69)	–4.72 (–5.95 to –3.50; 19.55)	<.001	
EM-introjected					.93
EBTT+OCTT	17.25 (3.40)	14.64 (2.83)	–2.61 (–4.35 to –0.86; 15.13)	.01	
OCTT×2	17.17 (4.48)	14.69 (3.39)	–2.48 (–4.53 to –0.43; 14.44)	.02	
EM–external regulation					<.001
EBTT+OCTT	24.54 (2.83)	23.11 (2.60)	–1.43 (–2.99 to 0.13; 5.83)	.07	
OCTT×2	24.34 (2.79)	18.24 (2.40)	–6.10 (–7.29 to –4.92; 25.06)	<.001	
Amotivation					.91
EBTT+OCTT	13.68 (2.98)	20.76 (2.06)	7.08 (6.31 to 9.04; 51.75)	.06	
OCTT×2	13.86 (3.16)	21.66 (2.44)	7.79 (6.20 to 9.38; 56.20)	<.001	

[a]MD: mean difference.
[b]IM: intrinsic motivation.
[c]EBTT+OCTT: exergame-based tennis training with on-court tennis training.
[d]OCTT: on-court tennis training.
[e]EM: extrinsic motivation.

Relationship Between Confidence Level and Tennis Technical Skill

For confidence, the bivariate correlation with tennis hit score (r=0.70; P<.001) and service speed (r=0.33; P=.01) was positive and significant. However, the correlation between confidence and service accuracy was not significant (r=0.21; P=.12). The coefficient of determination (r^2) for confidence with tennis hit score and service speed were 0.50 and 0.11, respectively.

Relationship Between Motivation Level and Tennis Technical Skill

IM had significant positive correlations with tennis hit score and service accuracy. The correlations with IM-to-know, IM-to-accomplish, and IM-to-experience stimulation were r=0.60, 0.45, and 0.80 for tennis hit score, and r=0.35, 0.40, and 0.40 for service accuracy. However, there was no significant correlation between tennis service speed score and IM-to-know, IM-to-accomplish, and IM-to-experience, where r=0.21 (P=.11), 0.18 (P=.18), and 0.18 (P=.18), respectively. The r^2 values for tennis hit score and IM-to-know, IM-to-accomplish, and IM-to-experience stimulation were 0.36, 0.21, and 0.62, respectively. For service accuracy, the r^2 values were 0.12, 0.16, and 0.16.

EM showed significant correlations (P<.001) only with tennis hit score, with r = 0.60 (EM-identified), r=0.67 (EM-introjected), and r=0.64 (EM–external regulation). There were no significant correlations with service speed ($P_{\text{EM-identified}}$=.21, $P_{\text{EM-introjected}}$=.72, and $P_{\text{EM–external regulation}}$=.65) or accuracy ($P_{\text{EM-identified}}$=.86, $P_{\text{EM-introjected}}$=.26, and $P_{\text{EM–external regulation}}$=.45). The r^2 values for tennis hit score and EM-identified, EM-introjected, and EM–external regulation were 0.36, 0.45, and 0.41, respectively.

Finally, amotivation was negatively correlated with tennis hit score (r=–0.39; P<.001), but there were no significant correlations with service speed (P=.46) or accuracy (P=.60). The r^2 between tennis hit score and amotivation was 0.15, meaning that 15% of the variability in tennis hit scores could be predicted by amotivation. Table 4 provides the correlation between grip strength, confidence, and motivation levels with tennis technical skills.

Table 4. Correlation analysis (Pearson r and 2-tailed *P* value) between grip strength, confidence, and motivation levels with tennis technical skills.

Outcome measures	Hit scores	Accuracy scores	Speed scores
Grip strength			
r	0.70[a]	0.45[a]	0.42[a]
P value	<.001	<.001	.001
Confidence level			
r	0.70[a]	0.21	0.33[a]
P value	<.001	.12	.01
IM[b]-to-know			
r	0.60[a]	0.35[a]	0.21
P value	<.001	.01	.11
IM-to-accomplish			
r	0.45[a]	0.40[a]	0.18
P value	<.001	<.001	.18
IM-to-experience stimulation			
r	0.80[a]	0.40[a]	0.18
P value	<.001	<.001	.18
EM[c]-identified			
r	0.60[a]	0.17	–0.02
P value	<.001	.21	.86
EM-introjected			
r	0.67[a]	0.05	–0.15
P value	<.001	.72	.26
EM–external regulation			
r	0.64[a]	0.06	–0.10
P value	<.001	.65	.45
Amotivation			
r	–0.39[a]	–0.07	–0.10
P value	<.001	.60	.46

[a]The correlation is significant at a significant level of .05 (2-tailed).

[b]IM: intrinsic motivation.

[c]EM: extrinsic motivation.

Discussion

Principal Findings

In general, both EBTT+OCTT and OCTT×2 groups demonstrated significant improvements (P<.001) in tennis technical skills, all with large effect sizes, though no significant between-group differences were found across hit scores (P=.97), service accuracy (P=.86), and service speed (P=.72). The absence of significant between-group differences should not be interpreted as evidence of equivalence between the interventions, as this study was not designed or powered to assess equivalence or noninferiority. Grip strength increased significantly (P<.001) in both groups, with no significant interaction effects (P=.24). Confidence levels also improved significantly (P<.001) in both groups, but without significant between-group differences (P=.31). In terms of motivation, the EBTT+OCTT group showed significant decreases (P<.001) in EM-identified and

EM-introjected, while the OCTT×2 group showed significant reductions ($P<.001$) in IM-to-know, IM-to-accomplish, EM-identified, and EM–external regulation. Notably, the OCTT×2 group experienced significantly greater decreases in IM-to-know, IM-to-accomplish, and EM–external regulation compared to EBTT+OCTT, all with large effect sizes. Correlational analysis revealed strong positive relationships ($r>0.50$) between hit skill and grip strength, confidence level, IM-to-know, IM-to-experience stimulation, and all EM subscales. However, a negative correlation was observed between hit skill and amotivation ($r=–0.39$).

Comparison With Previous Work

Effect of Exergaming on Grip Strength

This study found significant improvements in grip strength in both EBTT+OCTT and OCTT×2 groups. Grip strength increased by 1.5 times in the EBTT+OCTT group and by 1.6 times in the OCTT×2 group. These results align with previous research on sports training and muscular strength. The likely mechanism is the repeated high-intensity activation of the forearm and hand muscles required for racket control and stroke execution. A previous study and a review have corroborated that regular engagement in sports requiring sustained gripping actions and upper limb coordination, such as tennis, induces significant neuromuscular adaptations that manifest as measurable strength gains in distal upper extremity musculature [9,39]. The observed grip strength improvements in the EBTT+OCTT group (1.5 times) demonstrated comparability with findings from an electromyography biofeedback exergaming study, which reported a 0.17 times enhancement in grip strength following training [40]. The substantially greater improvements observed in this quasi-experimental study may be attributable to the longer intervention duration (12 weeks vs 2 weeks), allowing for more pronounced neuromuscular adaptation. These results support the theoretical framework suggesting that exergaming systems designed to replicate sport-specific movements can elicit muscular engagement patterns and strength development comparable to conventional training modalities [40]. While the between-group comparison revealed no statistically significant difference in grip strength improvement, the marginally lower gain in the EBTT+OCTT condition may reflect biomechanical differences in equipment characteristics. Specifically, the substantially lighter mass of the exergaming controller (90 g) compared to a standard tennis racket (280 g-300 g) potentially reduced the mechanical loading and strength demands during gameplay. Nevertheless, the significant improvements observed in the EBTT+OCTT group suggest that exergaming tennis, while not specifically designed for grip strengthening, can effectively enhance grip strength through repeated sport-specific movement patterns. These findings expand current understanding of EBTT+OCTT's therapeutic potential and its capacity to produce physiological adaptations similar to traditional training approaches.

Effect of Exergaming on Technical Skills

This study showed significant improvements in technical skill scores, with large effect sizes in both EBTT+OCTT and OCTT×2 groups. These results suggest that exergaming and traditional training are both effective for enhancing tennis skills. Improvements were likely driven by repetitive practice, whether in virtual or real settings, which aligns with previous research on motor learning in neurological rehabilitation [41]. This reflects the motor learning principle of practice, where skills must be rehearsed correctly to be mastered [42].

In the EBTT+OCTT group, hit scores increased by 5.5 times, service accuracy by 6.3 times, and service speed by 6.1 times. In the OCTT×2 group, hit scores rose by 4.2 times, service accuracy by 5.9 times, and service speed by 6.8 times. These findings are consistent with earlier studies, showing that exergaming can enhance motor skills and physical performance comparably to on-court training [42-45]. For example, one study reported that both exergaming and on-court training improved tennis players' reaction time by 1.03 times, indicating that exergaming may help players prepare for incoming shots [43]. Similar results have been observed in children, where exergaming improved forehand and backhand skills by 1.5 times compared to 1.33 times in traditional training [44]. The greater improvement observed in this study (average skill increase of 5.5 times) compared with previous work (around 1.3 times) may be explained by the novice status of participants. Before the training, their baseline skills were poor, scoring below 6 of 50 on the Hewitt tennis test. However, with structured guidance and consistent practice, their skills improved dramatically over 12 weeks. This reflects the principle of specificity, where training effects are specific to the practiced activity [46]. As EBTT+OCTT replicates tennis movements, skill gains in the virtual setting were transferred to real-world performance.

Supporting evidence comes from studies in other domains. Exergaming-based balance training improved balance by 1.67 times, similar to conventional balance training [47]. Another study on adolescent tennis players reported a 0.38-times increase in Universal Tennis Rating when training combined with virtual reality and tablet-based cognitive drills was compared to regular training [45]. Collectively, these findings highlight that exergaming can be as effective as traditional training for enhancing technical performance. The shared success likely stems from common motor learning principles and practice specificity [46,47]. Finally, this study extends the literature by focusing on novice players. Exergaming provided an engaging, space-efficient alternative for skill development, requiring only about 4×3 meters of training space.

Effect of Exergaming on Confidence Level

The study found that both EBTT+OCTT and OCTT×2 significantly improved participants' confidence in tennis performance. Confidence increased by 1.6 times in the EBTT+OCTT group and by 1.8 times in the OCTT×2 group. These results align with the self-efficacy theory, which highlights mastery experiences as a core source of confidence [48]. In this trial, structured and repetitive practice enabled participants to progressively master tennis skills, enhancing their belief in their abilities. The large effect sizes further indicate that both training types were highly effective in fostering a sense of control over performance, consistent with self-efficacy principles [49].

Previous research supports this view, showing that repeated practice in real sports settings increased athletes' confidence by 1.21 times, thus preparing them for competition [50]. Similarly, exergaming interventions have been shown to enhance confidence in other contexts, such as fall-prevention training for older adults, which raised confidence in avoiding falls by 1.05 times [51]. However, only a few studies have directly assessed exergaming's impact on sports confidence. This trial helps address that gap, showing that tennis-specific exergaming can enhance confidence comparably to on-court training. Notably, there was a study that found exergaming to be less effective than traditional training for psychological outcomes. The trial reported that while exergaming boosted enjoyment and engagement, it did not significantly improve confidence in novice athletes [52]. The difference may reflect variations in exergame type, intervention length, or participant skill level. This study used tennis-specific exergaming directly relevant to participants' skill goals, while the comparison study [52] used non–sport-specific games.

Overall, the comparable benefits of EBTT+OCTT and OCTT×2 suggest that exergaming can be a valuable addition to sports training, particularly where access to courts or equipment is limited. It may also provide a psychologically supportive environment for beginners or individuals with low self-confidence by reducing the pressures of real-world competition.

Effect of Exergaming on Motivation Level

In the EBTT+OCTT group, apart from an increase in amotivation levels, there was a decline in all other motivation categories. Significant decreases were observed only in IM-to-accomplish (by 0.09 times), EM-identified (by 0.22 times), and EM-introjected (by 0.15 times), with medium and small effect sizes, respectively. In the OCTT×2 group, aside from increases in amotivation and IM-to-experience, other motivation levels also declined. Significant decreases were noted in IM-to-know (by 0.28 times), IM-to-accomplish (by 0.24 times), EM-identified (by 0.20 times), and EM–external regulation (by 0.25 times), all with large effect sizes (ηp^2=0.12-0.30). These results suggest that both EBTT+OCTT and OCTT×2 may influence IM in complex ways. According to self-determination theory, IM is driven by the inherent satisfaction of performing an activity [53]. The observed decreases in IM components, such as IM-to-accomplish, could suggest that the initial novelty and enjoyment of mastering new skills may diminish over time as tasks become repetitive or as participants reach a plateau in skill development [53]. This aligns with past research showing that continuous, repetitive training can sometimes reduce IM, especially when the perceived challenge or novelty decreases [54]. This pattern is also supported by another study, where motivation level was reduced by 1.08 times in general, and amotivation increased by 3.69 times when the participants were provided with a wearable healthy lifestyle technology (Fitbit by Fitbit Inc) for 8 weeks [55]. These results may indicate that, over time, both intrinsic and EM decline. However, it should be noted that maintaining motivation is essential for sustaining motivation without the need to alter or add new training equipment. Thus, the results from this quasi-experimental study suggest that EBTT+OCTT was able to sustain IM more than OCTT×2.

The Interrelationship Between Grip Strength, Confidence Level, and Motivation Level Toward an Individual's Tennis Skills

The interaction between motivation, confidence, and grip strength appears central to enhancing tennis performance. In this study, grip strength showed strong positive correlations with tennis hit score (*r*=0.70), service accuracy (*r*=0.45), and service speed (*r*=0.42). This suggested that stronger grip strength contributed to more powerful and accurate strokes. Similar findings have been reported in previous studies, where grip-strengthening exercises improved forearm speed and service performance [10,20,56]. Stronger grip strength enhances the transfer of force from the lower body and core to the racket, leading to greater ball velocity and control. In this study, participants initially had low grip strength (mean=18.74 kg) and poor Hewitt's tennis test scores, often falling in the "D" or "F" range. Their early struggles in returning balls accurately highlighted the importance of grip strength for technical execution.

Exergaming, such as traditional training, improved grip strength in this study. The EBTT+OCTT group increased grip strength by 1.5 times, while the OCTT×2 group improved by 1.6 times. Although no significant difference was found between the groups, the repetitive upper-limb motions in exergaming likely contributed to strength gains. Past research supports this, showing that interactive games involving hand and arm actions can build muscular endurance and functional fitness [21,57,58]. The moderate correlation between grip strength and service speed (*r*=0.42) further supports its role in stroke performance. While this study did not establish equivalence between EBTT+OCTT and OCTT×2, the findings suggested that exergaming offered a meaningful, engaging way to strengthen tennis-related muscles.

Confidence also emerged as a critical factor. Posttraining, confidence levels rose significantly in both groups and were positively correlated with tennis hit scores (*r*=0.70) and service speed (*r*=0.33). Confidence helped participants take risks, handle pressure, and execute skills more effectively. Early in training, participants' low confidence coincided with poor accuracy and hesitancy. By the end of the training, confidence scores improved from 35 of 72 to 61 of 72, accompanied by fourfold increases in hit scores. Previous studies confirm that self-efficacy and confidence are linked to better performance, reduced anxiety, and resilience under stress [48,59-62]. Exergaming may support confidence development by providing progressive challenges and instant feedback, enabling athletes to see measurable progress. This study found confidence increased by 1.6 times in EBTT+OCTT and by 1.8 times in OCTT×2, showing that both methods are effective.

Motivation, particularly IM, also played a key role. Players with high IM were more committed to practice, strengthening both skills and confidence. In this study, IM correlated strongly with hit scores (*r*=0.60, 0.45, and 0.80 for IM-to-know, IM-to-accomplish, and IM-to-experience, respectively), while EM showed weaker links to performance. Although overall motivation decreased slightly over the 12 weeks, the decline was smaller in EBTT+OCTT, suggesting exergaming helped sustain interest. This aligns with previous research showing that task-oriented environments emphasizing skill mastery foster stronger IM than competition-focused settings [63-66]. Exergaming creates such an environment by offering clear goals, instant feedback, and playful competition, all of which help maintain engagement.

Taken together, the results suggest that grip strength, confidence, and motivation interact in a reinforcing cycle. Stronger grip strength enhances stroke performance, which boosts confidence. Increased confidence encourages effort and persistence, feeding back into skill gains. Intrinsic motivation sustains this process by making practice enjoyable and rewarding. Exergaming supports this loop by providing variety, feedback, and achievable challenges, which help maintain both confidence and motivation. While not a full replacement for on-court training, EBTT+OCTT appears to be a valuable complement, especially for novices or those with limited access to facilities. Exergaming requires minimal space, engages players psychologically, and offers cost-effective benefits in both sporting and rehabilitation contexts [49]. Coaches and practitioners could integrate exergaming to enhance early skill acquisition, sustain motivation, and reduce barriers to participation.

Exergaming Studies in Other Sports and Psychological Domains

The results of this study are consistent with previous exergaming interventions across a range of sports and populations. For example, in a study involving junior athletes practicing basketball, handball, and volleyball, exergaming significantly improved reaction time over 3 months, indicating its utility in sport-specific skill enhancement [67]. Similarly, in a soccer-based exergaming study, players demonstrated improved reaction time and passing accuracy, with larger effects observed in novices compared to experienced athletes [68]. These findings support the results of this quasi-experimental study, whereby novice players may be particularly responsive to exergaming interventions, potentially due to a greater scope for technical and psychological adaptation.

From a psychological perspective, the results align with previous studies showing that exergaming enhances motivation, confidence, and enjoyment. A study found that adolescents who engaged in exergaming reported significantly greater IM and self-efficacy compared to traditional physical education groups [69,70]. Likewise, another study demonstrated that virtual cycling games increased motivation and adherence among older adults [71]. While the effect sizes in this quasi-experimental study were large in terms of motivational outcomes, this may reflect the unique engagement and gamified feedback inherent in exergaming platforms, which are particularly effective in novice or recreational populations. In contrast, elite athletes may show more modest gains due to ceiling effects and previous high-level training.

Limitations

This study has several limitations that should be acknowledged. First, the findings may only be generalizable to novice players, as the participants recruited had less than 1 year of tennis experience. This may limit the applicability of the results to more experienced or elite players, who may respond differently to exergaming-based interventions. To address this, this quasi-experimental study standardized the training program and ensured participants had similar baseline experience levels. Future studies should explore the effects of exergaming among intermediate or advanced players to assess broader applicability.

Second, the dropout rate was approximately 12% in the OCTT×2 group and 15% in the EBTT+OCTT group. Although this attrition was within acceptable limits for a quasi-experimental design, it could have influenced group comparisons and reduced statistical power. Another key limitation was the group allocation switching that occurred before the start of the experiment, where 7 participants requested to join the group of their preference (OCTT×2 [n=3] and EBTT+OCTT [n=4]). This introduced self-selection bias and could potentially affect the internal validity of the study's results, especially for outcomes related to motivation. Furthermore, the per-protocol analysis of data reflected the partial allocation of participants' chosen group, as opposed to randomized groups. This limits the statistical strength and validity of the results reporting motivational outcomes typically associated with randomized designs. As such, findings about motivation in this study should be interpreted with caution, and future studies should consider more rigorous strategies to ensure the robustness of randomization. To mitigate this, the study recruited more participants than the minimum sample size required by power analysis to maintain sufficient statistical validity. Although participants who had requested the group change may have had pre-existing differences in motivation, their baseline levels remained similar compared to those who remained in their allocated groups. In addition, the consistency in results between the modified ITT and per-protocol analyses strengthens the validity of the findings, suggesting that the results are robust to missing data and protocol deviations. However, future studies should incorporate strategies such as participant incentives, participant blinding, or shorter program durations to enhance retention.

Third, the exergaming environment may not fully replicate the physical, cognitive, and environmental demands of real-life tennis. Factors such as court size, surface type, weather conditions, and opponent variability are absent in virtual simulations, potentially limiting the ecological validity of the findings. To address this limitation, EBTT+OCTT participants also received one on-court tennis session per week to complement the virtual training. However, further research should explore advanced platforms (for example, augmented or mixed reality) to better simulate real-world playing conditions.

Fourth, the sex distribution in the study sample was skewed, with approximately 63% female participants. This imbalance may introduce sex-related bias and limit the generalizability of the results across sexes. While the study analyzed group-level performance rather than sex-specific differences, future studies should aim for a more balanced male-to-female ratio or consider conducting sex-stratified analyses.

Fifth, although the study used validated tools to assess motivation, confidence, grip strength, and tennis skills, it relied primarily on quantitative methods. This limit understanding of participant experiences and perspectives regarding the training. Incorporating qualitative methods, such as interviews or focus groups, in future studies could provide richer insight into how and why exergaming may influence motivation and performance. At the same time, the assessor was not blinded to group allocation due to practical constraints inherent in sport-specific field testing, which could introduce potential bias. The tennis skill assessments required direct interaction and real-time scoring, making blinding operationally difficult and largely unfeasible. To minimize bias, a single trained assessor was used to ensure consistent scoring procedures across all participants.

Finally, while no statistically significant differences were found between EBTT+OCTT and OCTT×2 for most outcome measures, this should not be interpreted as evidence of equivalence between the 2 training modalities. The study used standard hypothesis testing and was not designed or powered to formally assess equivalence or noninferiority. Therefore, the finding of nonsignificant *P* values reflects a failure to reject the null hypothesis, not confirmation that the interventions are equally effective. Future studies using equivalence or noninferiority trial designs may be warranted to further evaluate whether EBTT+OCTT can produce outcomes comparable to traditional training.

Future Directions

Future research should consider several aspects to build upon the findings of this study. First, as mentioned previously, studies should explore the effectiveness of EBTT+OCTT in populations with varying levels of tennis proficiency, including intermediate and elite players, to determine whether skill transfer and psychological benefits persist across experience levels. Second, longitudinal studies with extended follow-up periods are required to assess the sustainability of improvements in technical skills, grip strength, motivation, and confidence beyond the 12-week intervention period. Third, more immersive technologies such as virtual reality or augmented reality could be integrated into exergaming systems to better replicate real-world conditions and improve ecological validity. Fourth, future trials should investigate how exergaming can be personalized based on individual learning styles, physical abilities, or motivational profiles, which may enhance engagement and outcomes. Fifth, incorporating qualitative methods could provide deeper insights into users' experiences and perceived barriers or facilitators to exergaming adoption. Finally, exergaming could be examined as a rehabilitation tool for individuals with musculoskeletal or neurological impairments, to determine its utility in clinical sports medicine or physiotherapy settings.

Conclusion

In conclusion, this quasi-experimental study demonstrated that both EBTT+OCTT and OCTT×2 significantly improved tennis technical skills, grip strength, and confidence level among novice players over 12 weeks. EBTT+OCTT also helped sustain participants' IM level. Although between-group differences were not significant for most outcomes, EBTT+OCTT was more effective in sustaining IM. The observed positive associations between tennis skills, grip strength, confidence, and motivation suggest a correlated process where improvements in one domain support gains in others. Overall, evidence points toward EBTT+OCTT as a feasible and engaging supplementary approach for developing both physical and psychological aspects of tennis performance in novice players. However, given the study's design and population, conclusions should be interpreted with caution and not generalized beyond novice recreational players. Further controlled studies are needed to confirm the broader applicability and long-term effects of exergaming in sport-specific training.

Acknowledgments

The content of the paper does not represent the opinion of the funder or the university itself. The funding sources did not influence any aspects of this study.

Funding

This research was funded by the Universiti Malaya Sustainable Development Centre grant LL2023ECO015 and the Universiti Malaya Faculty of Medicine Postgraduate Scholarship Scheme.

Data Availability

In accordance with the stipulations provided by the ethics governing committee (Universiti Malaya Ethics Committee), all data belongs to Universiti Malaya, and distribution of the data require previous permission from the committee. Study data can be provided upon reasonable request to the corresponding author.

Authors' Contributions

Conceptualization was led by JSN, with MMR and SZH contributing equally. Data curation and formal analysis were led by JSN, with MMR and SZH providing supporting contributions. Funding acquisition was led by MMR, with SZH contributing equally and JSN in a supporting role. Investigation, methodology, and project administration were led by JSN, with MMR and SZH providing support. Resources were managed by JSN (lead), with MMR and SZH contributing equally. Supervision was led by MMR, with SZH contributing equally and JSN supporting. Validation and visualization were led by JSN, with MMR and SZH

contributing equally. Writing of the original draft and review, and editing were led by JSN, with MMR and SZH providing supporting contributions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results before group reassignment for modified intention-to-treat analysis.
[DOCX File , 19 KB - games_v14i1e73732_app1.docx]

References

1. Sween J, Wallington S, Sheppard V, Taylor T, Llanos A, Adams-Campbell L. The role of exergaming in improving physical activity: a review. J Phys Act Health 2014;11(4):864-870 [FREE Full text] [doi: 10.1123/jpah.2011-0425] [Medline: 25078529]
2. Mat Rosly M, Mat Rosly H, Davis Oam GM, Husain R, Hasnan N. Exergaming for individuals with neurological disability: a systematic review. Disabil Rehabil 2017;39(8):727-735. [doi: 10.3109/09638288.2016.1161086] [Medline: 27108475]
3. Benzing V, Schmidt M. The effect of exergaming on executive functions in children with ADHD: a randomized clinical trial. Scand J Med Sci Sports 2019;29(8):1243-1253 [FREE Full text] [doi: 10.1111/sms.13446] [Medline: 31050851]
4. Guo W, Wang B, Smoter M, Yan J. Effects of open-skill exercises on cognition on community dwelling older adults: protocol of a randomized controlled trial. Brain Sci 2021;11(5):609 [FREE Full text] [doi: 10.3390/brainsci11050609] [Medline: 34068726]
5. Martinez-Millana A, Michalsen H, Berg V, Anke A, Gil Martinez S, Muzny M, et al. Motivating physical activity for individuals with intellectual disability through indoor bike cycling and exergaming. Int J Environ Res Public Health 2022;19(5):2914 [FREE Full text] [doi: 10.3390/ijerph19052914] [Medline: 35270607]
6. Khundam C, Nöel F. A study of physical fitness and enjoyment on virtual running for exergames. Int J Comput Games Technol 2021;2021:1-16. [doi: 10.1155/2021/6668280]
7. Levac DE, Huber ME, Sternad D. Learning and transfer of complex motor skills in virtual reality: a perspective review. J Neuroeng Rehabil 2019;16(1):121 [FREE Full text] [doi: 10.1186/s12984-019-0587-8] [Medline: 31627755]
8. Biddle SJH, Gorely T, Faulkner G, Mutrie N. Psychology of physical activity: a 30-year reflection on correlates, barriers, and theory. Int J Sport Exerc Psychol 2022;21(1):1-14. [doi: 10.1080/1612197x.2022.2147261]
9. Cronin J, Lawton T, Harris N, Kilding A, McMaster D. A brief review of handgrip strength and sport performance. J Strength Cond Res 2017;31(11):3187-3217. [doi: 10.1519/JSC.0000000000002149] [Medline: 28820854]
10. Kovač T, Marček T, Šarkanj B, Borišev I, Ižaković M, Jukić K, et al. Fullerol C60(OH)24 nanoparticles and drought impact on wheat (Triticum aestivum L.) during growth and infection with Aspergillus flavus. J Fungi (Basel) 2021;7(3):236 [FREE Full text] [doi: 10.3390/jof7030236] [Medline: 33809896]
11. Song Y, Lin H, Wang B, Yang M. Factors and measures affecting players' psychology in tennis. JSSHL 2024;7(3):106-111. [doi: 10.53469/jsshl.2024.07(03).14]
12. Hidayat Y, Yudiana Y, Hambali B, Sultoni K, Ustun UD, Singnoy C. The effect of the combined self-talk and mental imagery program on the badminton motor skills and self-confidence of youth beginner student-athletes. BMC Psychol 2023;11(1):35 [FREE Full text] [doi: 10.1186/s40359-023-01073-x] [Medline: 36737818]
13. Mercader-Rubio I, Ángel NG, Silva S, Furtado G, Brito-Costa S. Intrinsic motivation: knowledge, achievement, and experimentation in sports science students-relations with emotional intelligence. Behav Sci (Basel) 2023;13(7):589 [FREE Full text] [doi: 10.3390/bs13070589] [Medline: 37504036]
14. Bangari D, Choudhary PK, Choudhary S, Kandpal A, Singh H, Mohit. Effects of a 12-week integrated core and plyometric training program on tennis skills, agility, strength, and balance in adolescent tennis players. Pedagogy Phys Cult Sports 2025;29(4):308-319. [doi: 10.15561/26649837.2025.0408]
15. Buyle M, Jung Y, Pavlou M, Gonzalez S, Bamiou DE. The role of motivation factors in exergame interventions for fall prevention in older adults: a systematic review and meta-analysis. Front Neurol 2022;13:903673 [FREE Full text] [doi: 10.3389/fneur.2022.903673] [Medline: 35989930]
16. Birnstiel S, Morschheuser B. Exploring gamification in team sports: a scoping review and research agenda. 2024 Presented at: 8th International GamiFIN Conference 2024 (GamiFIN 2024); April 2-5, 2024; Ruka, Finland p. 33-46 URL: https://www.researchgate.net/publication/379873210_Exploring_gamification_in_team_sports_A_scoping_review_and_research_agenda
17. Biddiss E, Irwin J. Active video games to promote physical activity in children and youth: a systematic review. Arch Pediatr Adolesc Med 2010;164(7):664-672. [doi: 10.1001/archpediatrics.2010.104] [Medline: 20603468]
18. Qiu S, Kaisar E, Ding R, Han T, Hu J, Rauterberg M. Social balance ball: designing and evaluating an exergame that promotes social interaction between older and younger players. Int J Hum-Comput Interact 2023;40(11):2838-2861. [doi: 10.1080/10447318.2023.2175157]

19. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012;141(1):2-18. [doi: 10.1037/a0024338] [Medline: 21823805]
20. Genevois C, Frican B, Creveaux T, Hautier C, Rogowski I. Effects of two training protocols on the forehand drive performance in tennis. J Strength Cond Res 2013;27(3):677-682. [doi: 10.1519/JSC.0b013e31825c3290] [Medline: 22592176]
21. Sun H. Impact of exergames on physical activity and motivation in elementary school students: a follow-up study. J Sport Health Sci 2013;2(3):138-145. [doi: 10.1016/j.jshs.2013.02.003]
22. Sember V, Meh K, Sorić M, Starc G, Rocha P, Jurak G. Validity and reliability of international physical activity questionnaires for adults across EU Countries: systematic review and meta analysis. Int J Environ Res Public Health 2020;17(19):7161 [FREE Full text] [doi: 10.3390/ijerph17197161] [Medline: 33007880]
23. Barbosa DG, Vealey R, Pelegrini A, Farias G, Corazza S, Felden É. Validation and cultural adaptation of the sport confidence inventory (SCI) for Brazilian adolescent athletes. Rev Bras Psic Esporte 2022;11(2). [doi: 10.31501/rbpe.v11i2.12498]
24. Barreira J, Amaro AS, Borges K, Luque LL, Fernandes PT. Internal consistency of the sport motivation scale-II questionnaire in the Brazilian context: potentialities and limitations. Motriz: rev educ fis 2022;28:e10220010721. [doi: 10.1590/s1980-657420220010721]
25. Hewitt JE. Hewitt's tennis achievement test. Res Q Am Assoc Health Phys Educ 2013;37(2):231-240. [doi: 10.1080/10671188.1966.10613366]
26. Reuter SE, Massy-Westropp N, Evans AM. Reliability and validity of indices of hand-grip strength and endurance. Aust Occup Ther J 2011;58(2):82-87. [doi: 10.1111/j.1440-1630.2010.00888.x] [Medline: 21418230]
27. Eugster N, Ewerhart C. Rules in international tennis tournaments: a game theoretic perspective. J Sport Game 2024;6(1):33-42. [doi: 10.22259/2642-8466.0601004]
28. Pescatello L. ACSM's Guidelines for Exercise Testing and Prescription. Philadelphia, PA: Lippincott Williams & Wilkins; 2014.
29. Sawali L, Refiater UH. Effectiveness of training programs to improve tennis service skills. HON 2022;5(2):609. [doi: 10.31851/hon.v5i2.8279]
30. González-Rosalén J, Benítez-Martínez JC, Medina-Mirapeix F, Cuerda-Del Pino A, Cervelló A, Martín-San Agustín R. Intra- and inter-rater reliability of strength measurements using a pull hand-held dynamometer fixed to the examiner's body and comparison with push dynamometry. Diagnostics (Basel) 2021;11(7):1230 [FREE Full text] [doi: 10.3390/diagnostics11071230] [Medline: 34359313]
31. Mathiowetz V. Comparison of Rolyan and Jamar dynamometers for measuring grip strength. Occup Ther Int 2002;9(3):201-209 [FREE Full text] [doi: 10.1002/oti.165] [Medline: 12374997]
32. Jaafar MH, Ismail R, Ismail NH, Md Isa Z, Mohd Tamil A, Mat Nasir N, et al. Normative reference values and predicting factors of handgrip strength for dominant and non-dominant hands among healthy Malay adults in Malaysia. BMC Musculoskelet Disord 2023;24(1):74 [FREE Full text] [doi: 10.1186/s12891-023-06181-8] [Medline: 36709276]
33. Kamarul T, Ahmad T, Loh W. Hand grip strength in the adult Malaysian population. J Orthop Surg (Hong Kong) 2006;14(2):172-177 [FREE Full text] [doi: 10.1177/230949900601400213] [Medline: 16914783]
34. Pelletier LG, Tuson KM, Fortier MS, Vallerand RJ, Briére NM, Blais MR. Toward a new measure of intrinsic motivation, extrinsic motivation, and amotivation in sports: the sport motivation scale (SMS). J Sport Exerc Psychol 1995;17(1):53. [doi: 10.1123/jsep.17.1.35]
35. Vealey RS. Conceptualization of sport-confidence and competitive orientation: preliminary investigation and instrument development J. Sport Psychol 1986;8(3):221-246 [FREE Full text] [doi: 10.1123/jsp.8.3.221]
36. Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Ann Card Anaesth 2019;22(1):67-72 [FREE Full text] [doi: 10.4103/aca.ACA_157_18] [Medline: 30648682]
37. Richardson JTE. Eta squared and partial eta squared as measures of effect size in educational research. Educ Res Rev 2011;6(2):135-147. [doi: 10.1016/j.edurev.2010.12.001]
38. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge; Dec 1988.
39. Meigh NJ, Keogh JWL, Schram B, Hing W, Rathbone EN. Effects of supervised high-intensity hardstyle kettlebell training on grip strength and health-related physical fitness in insufficiently active older adults: the BELL pragmatic controlled trial. BMC Geriatr 2022;22(1):354 [FREE Full text] [doi: 10.1186/s12877-022-02958-z] [Medline: 35459114]
40. Garcia-Hernandez N, Garza-Martinez K, Parra-Vega V. Electromyography biofeedback exergames to enhance grip strength and motivation. Games Health J 2018;7(1):75-82. [doi: 10.1089/g4h.2017.0054] [Medline: 29227162]
41. Levin MF, Demers M. Motor learning in neurological rehabilitation. Disabil Rehabil 2021;43(24):3445-3453. [doi: 10.1080/09638288.2020.1752317] [Medline: 32320305]
42. Kitago T, Krakauer JW. Chapter 8 - motor learning principles for neurorehabilitation. In: Handbook of Clinical Neurology. Amsterdam, NL: Elsevier; 2013:93-103.
43. Flôres F, Silva A, Willig R, Reyes A, Serpa J, Marconcin P, et al. Exergames as an effective alternative to real environmental tennis training for adolescents. Sports (Basel) 2024;12(6):168 [FREE Full text] [doi: 10.3390/sports12060168] [Medline: 38921862]

44. Šlosar L, de Bruin ED, Fontes EB, Plevnik M, Pisot R, Simunic B, et al. Additional exergames to regular tennis training improves cognitive-motor functions of children but may temporarily affect tennis technique: a single-blind randomized controlled trial. Front Psychol 2021;12:611382 [FREE Full text] [doi: 10.3389/fpsyg.2021.611382]
45. Anguera JA, Choudhry A, Seaman M, Fedele D. Assessing the impact of a virtual reality cognitive intervention on tennis performance in junior tennis players: pilot study. JMIR Form Res 2025;9:e66979 [FREE Full text] [doi: 10.2196/66979] [Medline: 39918854]
46. Barnett ML, Ross D, Schmidt RA, Todd B. Motor skills learning and the specificity of training principle. Res Q Am Assoc Health Phys Educ 2013;44(4):440-447. [doi: 10.1080/10671188.1973.10615224]
47. Gioftsidou A, Vernadakis N, Malliou P, Batzios S, Sofokleous P, Antoniou P, et al. Typical balance exercises or exergames for balance improvement? J Back Musculoskelet Rehabil 2013;26(3):299-305. [doi: 10.3233/BMR-130384] [Medline: 23893145]
48. Feltz DL. Self-confidence and sports performance. Exerc Sport Sci Rev 1988;16(1):423-458. [doi: 10.1249/00003677-198800160-00016]
49. Bandura A. Self-Efficacy: The Exercise of Control. New York, NY: WH Freeman/Times Books/Henry Holt & Co; 1997.
50. Heydari A, Soltani H, Mohammadi-Nezhad M. The effect of psychological skills training (goal setting, positive selftalk and Imagery) on self-confidence of adolescent volleyball players. Pedagog psychol med-biol probl phys train sports 2018;22(4):189. [doi: 10.15561/18189172.2018.0404]
51. Stanmore EK, Mavroeidi A, de Jong LD, Skelton DA, Sutton CJ, Benedetto V, et al. The effectiveness and cost-effectiveness of strength and balance Exergames to reduce falls risk for people aged 55 years and older in UK assisted living facilities: a multi-centre, cluster randomised controlled trial. BMC Med 2019;17(1):49 [FREE Full text] [doi: 10.1186/s12916-019-1278-9] [Medline: 30813926]
52. McGonigal J. Reality is broken: why games make us better and how they can change the world. Choice Reviews Online 2011;49(02):49-0680. [doi: 10.5860/choice.49-0680]
53. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78 [FREE Full text] [doi: 10.1037//0003-066x.55.1.68]
54. Grouzet FME, Vallerand RJ, Thill EE, Provencher PJ. From environmental factors to outcomes: a test of an integrated motivational sequence. Motiv Emot 2004;28(4):331-346. [doi: 10.1007/s11031-004-2387-z]
55. Kerner C, Goodyear VA. The motivational impact of wearable healthy lifestyle technologies: a self-determination perspective on fitbits with adolescents. Am J Health Educ 2017;48(5):287-297. [doi: 10.1080/19325037.2017.1343161]
56. Nugroho D, Hidayatullah MF, Doewes M, Purnama SK. Does grip strength, performance, and hand-eye coordination affect tennis drive skills? In: Proceedings of the International Conference on Learning and Advanced Education. 2023 Presented at: ICOLAE 2022; December 6; Surakarta, Indonesia p. 4-11. [doi: 10.2991/978-2-38476-086-2_2]
57. Zeng N, Pope Z, Lee J, Gao Z. Virtual reality exercise for anxiety and depression: a preliminary review of current research in an emerging field. J Clin Med 2018;7(3):42 [FREE Full text] [doi: 10.3390/jcm7030042] [Medline: 29510528]
58. Lin SC, Lee JY, Yang Y, Fang CC, Fang HL, Hou TH. Exploring the design of upper limb strength training through high-intensity interval training combined with exergaming: usability study. JMIR Serious Games 2024;12:e51730 [FREE Full text] [doi: 10.2196/51730] [Medline: 38632713]
59. Covassin T, Pero S. The relationship between self-confidence, mood state, and anxiety among collegiate tennis players. J Sport Behav Sep 2004;27(3):230 [FREE Full text]
60. Kajtna T, Barić R. Psychological characteristics of coaches of successful and less successful athletes in team and individual sports. Rev. Psychol 2009;16(1):47-56 [FREE Full text]
61. Gomes RV, Cunha VCR, Zourdos MC, Aoki MS, Moreira A, Fernandez-Fernandez J, et al. Physiological responses of young tennis players to training drills and simulated match play. J Strength Cond Res 2016;30(3):851-858. [doi: 10.1519/JSC.0000000000001159] [Medline: 26382129]
62. Hahn E. Emotions in sports. In: Anxiety in Sports. Abingdon, United Kingdom: Taylor & Francis; 2021.
63. Chen SP, Shang Y, Xie LJ, Song D, Chen M. The true value of extrinsic motivation? The effect of intrinsic and extrinsic sports motivation on sports anomie behavior in college students. Curr Psychol 2023;43(3):2704-2713. [doi: 10.1007/s12144-023-04524-1]
64. Tuckey M, Brewer N, Williamson P. The influence of motives and goal orientation on feedback seeking. J Occup Organ Psychol 2010;75(2):195-216. [doi: 10.1348/096317902600986771]
65. Cervelló E, Santos Rosa FJ, Calvo TG, Jiménez R, Iglesias D. Young tennis players' competitive task involvement and performance: the role of goal orientations, contextual motivational climate, and coach-initiated motivational climate. J Appl Sport Psychol 2007;19(3):304-321. [doi: 10.1080/10413200701329134]
66. Thibodeaux J, Winsler A. Stay positive: self-talk in relation to motivational climate, goal orientation, and self-talk encouragement in tennis. Res Q Exerc Sport 2022;93(4):769-780. [doi: 10.1080/02701367.2021.1918323] [Medline: 34723765]
67. Badau D, Badau A, Ene-Voiculescu C, Larion A, Ene-Voiculescu V, Mihaila I, et al. The impact of implementing an exergame program on the level of reaction time optimization in handball, volleyball, and basketball players. Int J Environ Res Public Health 2022;19(9):5598 [FREE Full text] [doi: 10.3390/ijerph19095598] [Medline: 35564993]

68. Bideau B, Kulpa R, Vignais N, Brault S, Multon F, Craig C. Using virtual reality to analyze sports performance. IEEE Comput Graph Appl 2010;30(2):14-21. [doi: 10.1109/MCG.2009.134] [Medline: 20650707]
69. Staiano AE, Calvert SL. Exergames for physical education courses: physical, social, and cognitive benefits. Child Dev Perspect 2011;5(2):93-98 [FREE Full text] [doi: 10.1111/j.1750-8606.2011.00162.x] [Medline: 22563349]
70. Ochi G, Kuwamizu R, Fujimoto T, Ikarashi K, Yamashiro K, Sato D. The effects of acute virtual reality exergaming on mood and executive function: exploratory crossover trial. JMIR Serious Games 2022;10(3):e38200 [FREE Full text] [doi: 10.2196/38200] [Medline: 36169992]
71. Anderson-Hanley C, Arciero PJ, Brickman AM, Nimon JP, Okuma N, Westen SC, et al. Exergaming and older adult cognition: a cluster randomized clinical trial. Am J Prev Med 2012;42(2):109-119. [doi: 10.1016/j.amepre.2011.10.016] [Medline: 22261206]

Abbreviations

EBTT+OCTT: exergame-based tennis training with on-court tennis training
EM: extrinsic motivation
IM: intrinsic motivation
IPAQ: International Physical Activity Questionnaire
ITT: intention-to-treat
MD: mean difference
OCTT: on-court tennis training
SCI: Sports Confidence Inventory
SMT: Sports Motivation Test

Edited by A Coristine; submitted 17.Mar.2025; peer-reviewed by BCH Huijgen, N Deng; comments to author 23.May.2025; revised version received 15.Dec.2025; accepted 17.Dec.2025; published 02.Feb.2026.

Please cite as:
Ngo JS, Hoe SZ, Mat Rosly M
Effects of Exergaming Tennis on Players' Tennis Skills and Mental State Compared to Regular Tennis in Adult Players: Quasi-Experimental Study
JMIR Serious Games 2026;14:e73732
URL: https://games.jmir.org/2026/1/e73732
doi:10.2196/73732
PMID:

Immersive Tai Chi for Home-Based Exercise in Older Adults: Usability and Feasibility Study

XiaCheng Song[1], PhD; Nazlena Mohamad Ali[1], PhD; Mohamad Hidir Mhd Salim[1], PhD; Muhammad Yudhi Rezaldi[2,3], PhD

[1]Institute of Visual Informatics, National University of Malaysia, Aras 1, Bangi, Selangor, Malaysia
[2]Research Centre for Data and Information Sciences, National Research and Innovation Agency, Bandung, Indonesia
[3]Faculty of Creative Industry, Telkom University, Bandung, Indonesia

Corresponding Author:
Nazlena Mohamad Ali, PhD
Institute of Visual Informatics, National University of Malaysia, Aras 1, Bangi, Selangor, Malaysia

Abstract

Background: Longer life expectancy makes physical exercise crucial for active aging. However, adherence to traditional exercise among community-dwelling older adults is generally low. Virtual reality (VR) and mixed reality (MR) Tai Chi exergames, as novel health promotion tools, show significant potential, particularly for older adults exercising in a home setting.

Objective: This study aimed to evaluate the usability and feasibility of a VR and MR Tai Chi exergame for community-dwelling older adults, focusing on subjective experience, physiological comfort, and objective interaction performance. The study also explored the relationships between key usability factors and sought to quantify links between objective accuracy and subjective experience (the Game Experience Questionnaire or Virtual Reality Sickness Questionnaire [VRSQ]) to inform choices of display mode, feedback strength, and session length.

Methods: Of the 86 community-dwelling older adults recruited for this study, data from 70 participants were considered valid after an initial screening, during which 16 (18.6%) were excluded due to issues with VR adaptation. Participants were sequentially assigned in a rotating order to 1 of 4 variants (VR/MR×soothing/intense) to balance exposure; however, primary analyses were preplanned to be collapsed across variants, focusing on whole-sample usability and feasibility rather than confirmatory between-group hypotheses. The primary outcome measures included the Game Experience Questionnaire, VRSQ, and objective gameplay logs.

Results: The VR/MR Tai Chi game demonstrated good overall usability and acceptability among the screened participants. Subjective experience was highly positive, with median scores for "positive affect" (median 4.0) and "competence" (median 3.8) being significantly high, whereas the median for "Challenge" (median 1.4) was significantly low ($P<.001$ for all). Physiological comfort in the postscreening sample was acceptable, with the most common mild symptoms being dizziness with eyes closed (20.0%) and vertigo (18.6%), both of low severity; however, the initial exclusion of 18.6% of participants due to VR discomfort is noteworthy. Therefore, generalizability is limited because the analyzed sample overrepresents older adults who tolerate immersive displays. Accuracy showed significant positive correlations with flow (ρ=0.342) and competence (ρ=0.322), whereas the VRSQ total score was significantly negatively correlated with positive affect (ρ=−0.334, $P=.005$).

Conclusions: Tai Chi exergames based on immersive technologies offer a feasible and attractive pathway for promoting physical exercise among community-dwelling older adults, particularly within the home environment, supporting the goal of aging in place. As a single, laboratory-based session, the reported satisfaction may partly reflect a novelty effect; therefore, longer-term, home-based follow-ups are needed to assess durability. Analysis of the key usability factors provides guidance for specific design choices, while also indicating directions for future research, such as longitudinal evaluations, extension to more diverse populations, and application in real-world home settings.

(JMIR Serious Games 2026;14:e79453) doi:10.2196/79453

KEYWORDS

virtual reality; video games; aged; Tai Chi; usability testing; motion sickness

Introduction

The trend of people getting older around the world is accelerating. By 2050, there will be more than 2 billion people over the age of 60 years, and older adults will make up a larger share of the world's population, especially in developing areas [1]. This change in the population puts more pressure on the economy and shows how important programs such as "Healthy

Aging" and letting older people "age in place" are. Regular exercise is very important for healthy aging [2]. It helps keep people healthy, helps them stay independent, makes them feel better mentally, and improves their quality of life [3]. These benefits make exercise a good and useful way to help older people all over the world. However, older adults often face numerous barriers that make it hard for them to participate in traditional exercise programs. These problems include not being motivated, being afraid of falling, and not being able to get to the right facilities [4]. These challenges show that conventional exercise models often fail to meet the emotional and experiential needs of older adults. Recent studies suggest that digitally delivered, personalized interventions, such as programs incorporating wearable technology, remote coaching, and motivational messaging, can significantly improve adherence and reduce fear of falling [5,6]. This points to the need for approaches that improve enjoyment and intrinsic motivation in older adults.

To tackle this issue, immersive exergames that combine virtual reality (VR) and mixed reality (MR) technologies are a viable solution. This research concentrates on the integration of Tai Chi with VR/MR technologies. Tai Chi is particularly suitable for older adults due to its proven efficacy in enhancing dynamic balance and substantially decreasing both the frequency of falls and the fear of falling, which is a significant psychological impediment to physical activity engagement in this demographic [7]. These advantages render Tai Chi a pragmatic and psychologically beneficial approach for enhancing mobility and self-assurance in older adults. VR/MR Tai Chi exergames are meant to be a safe, easy, and fun way to work out at home. These exergames can address many limitations of traditional exercise by providing virtual practice environments that are controllable. They may also boost intrinsic motivation by enhancing feelings of autonomy and competence. Although extended reality technologies hold substantial promise for helping older adults, several challenges remain, such as the high cost of equipment, steep learning curves, and cybersickness. Notably, previous research has emphasized that cybersickness is a frequent and underestimated barrier that may lead to user dropout and restrict broader adoption in unsupervised home settings [8,9]. These usability and safety issues should be addressed before extended reality solutions can be widely implemented among older populations.

Although VR and MR exergames show promise for helping older adults avoid falling, there is insufficient systematic research comparing how different display modes and feedback systems affect this group. Some early evidence suggests that augmented reality, a type of MR, may be better for older adults because it gives them a better sense of their body and space, which makes them feel less likely to fall [10]. This points to a research gap in comprehending the preferences and reactions of older users to various immersive environments. Recognizing the distinct interactive experiences of older adults is essential for effective design, particularly in unsupervised home contexts. Hence, distinguishing between VR and MR modes goes beyond being a mere technical matter; it has direct implications for the safety, comfort, and acceptance of these technologies among older adults using them independently at home.

This work provides evidence to guide the development of immersive health care applications to enable older individuals to practice independently at home, a key component within the "aging in place" plan. With professional guidance lacking within home contexts, intrinsic motivational abilities and efficacy and safety of the game within itself become paramount [11]. Thus, this work's investigation into aspects such as cybersickness and "accuracy-competence paradox" remains a key focal point within this work to aid in determining if such technology can be disseminated and leveraged effectively.

Specifically, this study will seek to answer the following core questions:

1. Overall usability and acceptance: How does age affect subjective experience (eg, "empowerment" feelings), physiological comfort (ie, cybersickness symptom occurrence), and objective interaction efficacy (eg, the "accuracy-competence paradox") while playing the VR/MR Tai Chi game?
2. Internal links between fundamental factors: How does objective gameplay performance (eg, pointing precision) influence subjective psychological experiences (eg, flow, competence)? Do users' physiological symptoms of discomfort affect positive emotions? Are positive emotions correlated to users' inherent traits, such as age?

Through systematic investigation of these questions, this study aimed to contribute empirical evidence to be used in designing immersive health applications among older people [12].

Methods

Ethical Considerations

The study received ethics approval and adhered to guidelines for research involving human participants.

Research Ethics Approval

This study received ethics approval from the Institute of Visual Informatics, Universiti Kebangsaan Malaysia. The approval was granted on July 2, 2024, with the reference number UKM.IVI.600-4/6/P130610.

Informed Consent

All participants provided written informed consent before the start of any experimental procedures. The researchers clarified the study's objectives, methods, possible risks (including cybersickness symptoms), and potential benefits. They also made it clear to participants that they could withdraw from the study at any point without facing any consequences.

Privacy and Confidentiality

All personally identifiable information and harvested data were fully anonymized to ensure respondent privacy. Data were stored securely on a password-protected server, and only central research team staff had access.

Compensation

Participants did not receive any monetary or nonmonetary compensation for their participation. Their involvement in this study was entirely voluntary.

Study Design

Overview

This study was designed as a pilot study using a 2×2 factorial design (display mode: VR/MR×feedback: soothing/intense) to effectively explore the feasibility of varying exergame configurations before a large trial. Participants were sequentially assigned to 1 of the 4 conditions using a rotating 1:1:1:1 schedule upon enrollment. The primary goal of this multicondition design was not to conduct formal hypothesis testing between groups, which would require a much larger sample. Instead, its purpose was to ensure that our usability and feasibility assessment covered a range of potential interaction styles. This approach allows us to identify any major negative outcomes (eg, excessive cybersickness or frustration) that might be unique to a specific condition, providing critical insights for optimizing a single, refined protocol for future, larger-scale home-based studies.

Recruitment, Screening, and Final Sample

A research team initially recruited 86 community-dwelling older people who were capable of living independently, via media such as community promotions at community centers and newspaper advertising locally. Such a process was used to maintain the sample's characteristics closely aligned with those of the target end user group for future home-based VR/MR Tai Chi exergames [13].

In this pilot phase of research, screening led to a dropout of 16 participants (16/86, 18.6% of the original sample) who could not accommodate to the virtual environment or who experienced severe dizziness. Notably, this screening process suggests that the final sample was skewed to some degree to VR tolerance, that is, comprised individuals better predisposed to accommodate virtual environment exposures. This characteristic is relevant in estimating the extent to which the findings generalize to a broader community of older individuals.

We performed an analysis of the 16 participants who were excluded due to a failure to adapt to VR. This group comprised 11 (68.8%) women and 5 (31.2%) men, a gender distribution similar to that of the retained sample. However, a systematic difference was evident in age: 43.8% (7/16) of participants excluded were aged older than 75 years, a proportion more than double that of the sample maintained (15/70, 21.4%). This finding suggests that older age could be a determining factor in intolerance to VR in this population.

Inclusion and Exclusion Criteria

The aforementioned recruitment and screening processes adhered to clearly defined criteria. The inclusion criterion was enrollment of participants aged 60 years and older [14,15]. Exclusion criteria encompassed any known adverse reactions to VR experiences and severe cardiorespiratory or musculoskeletal diseases that would significantly impede participation in the experimental procedures, severe vision loss, and other health conditions unsuitable for the use of immersive devices (eg, vertigo and epilepsy).

Self-report alone was considered insufficient; therefore, a brief experiential screening was recommended before unsupervised use.

Demographic Characteristics, Health Status, and Exercise Habits

Data regarding demographic information, health background, and daily exercise habits were collected from the 70 valid participants using a self-developed "basic information questionnaire." The detailed characteristics are presented in Table 1.

The 70 valid participants included in the final analysis exhibited certain demographic tendencies: they were predominantly female (52/70, 74.29%) and largely comprised "younger" older adults aged 60 to 75 years (55/70, collectively 78.56%). In terms of health status, although a considerable proportion (40/70, 57.14%) self-reported chronic conditions such as cardiopulmonary diseases, they were also individuals who actively managed their health, with a high percentage (66/70, 94.29%) undergoing regular physical examinations. This characteristic of "living actively with illness" suggests that the sample group might hold a more open attitude toward health-promoting interventions, but it also necessitates that such interventions ensure safety and appropriateness.

Table . Participant demographic characteristics, health status, and exercise habits (N=70).

Characteristics	Values, n (%)
Sex	
Male	18 (26)
Female	52 (74)
Age distribution (y)	
60 - 65	32 (46)
65 - 70	11 (16)
71 - 75	12 (17)
75 - 80	11 (16)
>80	4 (6)
Height distribution (cm)	
140 - 150	13 (19)
150 - 160	36 (51)
160 - 170	20 (29)
>170	1 (1)
Weight distribution (kg)	
30 - 40	2 (3)
40 - 50	8 (11)
50 - 60	36 (51)
60 - 70	18 (26)
70 - 80	4 (6)
>80	2 (3)
Self-reported illness status	
No reported cardiopulmonary illness	30 (43)
With cardiopulmonary disease	40 (57)
Regular physical examination	
Yes	66 (94)
No	4 (6)
Weekly exercise frequency	
Exercise daily	66 (94)
Exercise 3 - 5 times per week	1 (1)
Rarely exercise	3 (4)

This pilot group reported very high levels of physical activity, such that the vast majority (66/70, 94.29%) used a daily exercise regimen and participated in a broad range of activities, such as walking, square dancing, and Tai Chi. Overall, this sample is a subgroup of predominantly female, aged between 60 and 75 years, who exercise on a regular basis, approach health care management in an active manner, and have good tolerance to VR environment (as a consequence of preliminary screening that eliminated 18.6% (16/86) who reported dizziness while experiencing VR). These characteristics are relevant to explain why this sample reports such a high level of interest and positive response to the VR Tai Chi game. While these same characteristics will limit how widely the investigation can be generalized to broader older groups, such as those who exercise less or who possess poorer VR tolerance.

VR Tai Chi Exergame and Experimental Equipment

Intervention and Equipment

The VR Tai Chi exergame used in this study was designed for older individuals. Its fundamentals can be equally analyzed using the MDA (mechanics, dynamics, and aesthetics) approach [16,17].

Mechanics

It guides users through 16 selected Tai Chi exercises. It is interactive via a head-mounted display (HMD) and hand controllers reacting to head and hand movements. Users must synchronize their pose to 3 hotspots (head and both hands) provided in the on-screen instruction. A movement only proceeds after a successful success window, and a movement can save up to 3 hotspot holds (left hand, right hand, and HMD) before advancing. If users complete the entire sequence of 16 movements ahead of schedule, they will repeat the sequence immediately, beginning with the first movement. This generous success condition was a conscious design choice intended to provide a sense of accomplishment. It has 2 display modes, VR and MR, and can come equipped with either soothing or upbeat music and vibration feedback. Additionally, players can skip the current move or pause at will during gameplay.

Dynamics

Playing involves players' efforts at mirroring exemplified postures, receiving immediate multisensory feedback, and regulating progression during a trial lasting a maximum of 5 minutes. One key dynamic reveals itself in the gamification mechanics: participants can choose to meet only the bare minimum threshold needed for progress ("single hotspot for 3 s") to progress effortlessly or shoot for immaculate matching across all hotspots to progress to a higher "accuracy" level. This allows participants to balance challenge against victory on a preference and ability level. It is this dynamic that later results regarding user experience build upon.

Aesthetics

The game aims to evoke multidimensional aesthetic experiences, such as a "sense of competence" from successfully completing movements, a "sense of challenge" from improving performance, feelings of "relaxation or motivation" engendered by different musical feedback, and a "sense of immersion" provided by the VR/MR environments. These subjective experiences are intended to encourage physical exercise in older users in an engaging manner, and their achievement was assessed using the Game Experience Questionnaire (GEQ) [18,19]. Design philosophy appears to be interested in desirable affective effects and perceived competence corresponding to self-determination theory principles, such that satisfying the need for competence can improve intrinsic motivation.

Experimental Equipment

This study used the Meta Quest 3 (Meta Platforms, Inc) HMD for presenting both VR and MR conditions [20,21]. The choice of specific, commercially available hardware, such as the Meta Quest 3, enhances the transparency and potential replicability of the study.

Activity Area

Exercise took place in an area of approximately 5 m^2, comparable to a typical living room, to ensure home feasibility and safety [22,23].

Experimental Procedure

Group allocation followed a prespecified rotating schedule (1:1:1:1) prepared before data collection; participants had no influence on assignment.

Experimental Preparation and Adaptation Training

When participants arrived at the laboratory, they first completed an informed consent process and a general information questionnaire. Next, a short adaptation training session was completed to enable participants (and especially those without any training in VR/MR) to become familiar with both the immersive environment and how to use the equipment. This adaptation step is a valuable step in acclimating older adults to new technology, such as VR, because it can be used to help reduce technology anxiety and achieve a minimum baseline level of comfort and competence in how to use the equipment before beginning the intervention [24,25].

Exergame Intervention

A visual summary of the participant flow and 2×2 design (VR/MR×soothing/intense) appears in Figure 1. This approach ensured that group sizes remained balanced and that potential confounding factors, such as personal preference or prior experience, were minimized. Participants followed a fixed 16-movement sequence; each 5-minute session could end before all movements were finished due to the time limit or restart at the first movement if the sequence was completed early. After this orientation, each participant performed 2 formal 5-minute gameplay sessions under their assigned experimental condition (VR/MR×soothing/intense feedback). The better performance out of the 2 sessions was recorded for analysis. All single-session metrics (eg, posture, accuracy) were computed per session; the higher of the 2 sessions was retained for analysis. This procedure ensured that every participant experienced the complete set of movements while also minimizing the impact of first-trial unfamiliarity on the final outcome [26].

VR provided a fully virtual scene; MR overlaid guidance onto the real room (Figure 2). "Soothing" used gentle cues; "intense" used more salient multimodal cues; task rules were identical [27].

Figure 1. Participant Flow and 2×2 Study Design. MR: mixed reality; VR: virtual reality.

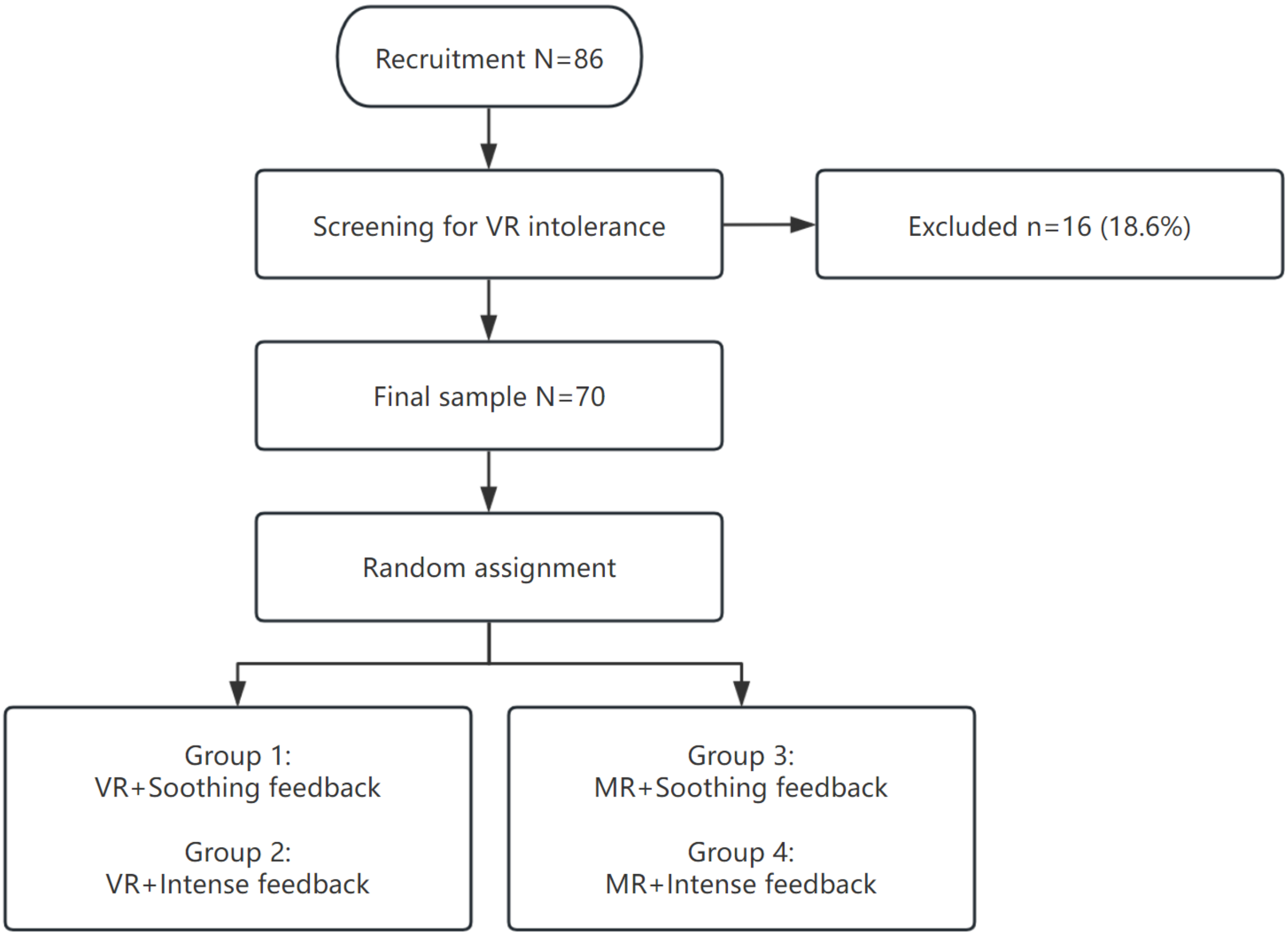
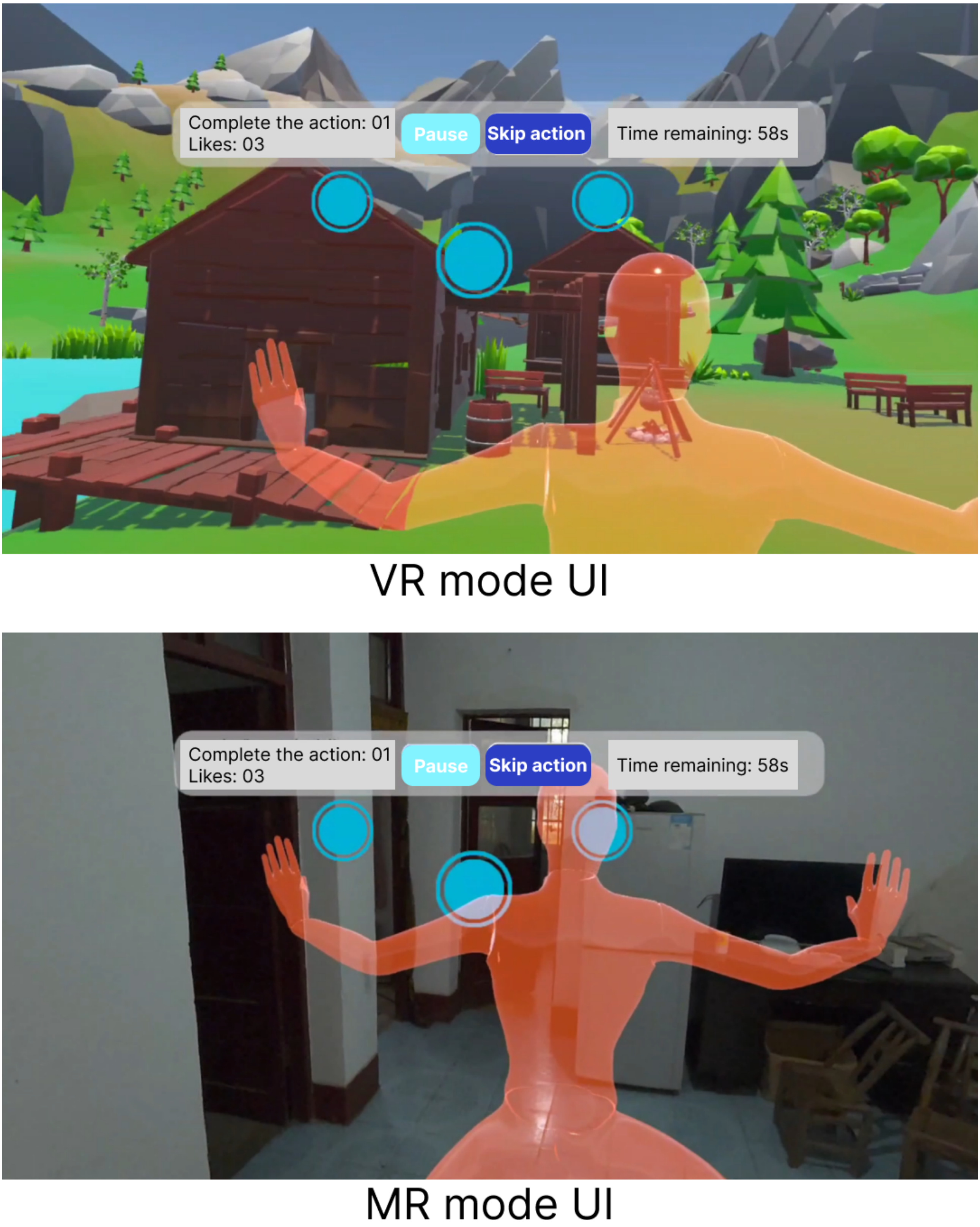


Figure 2. Screenshots of the VR and MR display modes. MR: mixed reality; UI: user interface; VR: virtual reality.



Safety and Supervision

The experimental protocol was observed and supervised in a one-on-one capacity by trained researchers to be able to control and respond to any new sensory stimulus or sensation of disorientation. Subjects were informed they could withdraw from participation at any time if they experienced any type of discomfort or abnormal level of fatigability. Such precautions are particularly relevant in research among older individuals and potentially disorienting technologies such as VR. To regulate fatigue and potential dizziness between sessions, exposure was scheduled in 2 approximately 5-minute sessions, and participants could rest and pause whenever needed.

This was a single-visit, supervised laboratory study, consisting of the 2 gameplay blocks described earlier, with no home-based or longitudinal follow-up. Postsession questionnaires captured immediate responses; repeated exposures to control for novelty were not implemented.

Outcome Measures

Overview

This research took a multidimensional perspective to assess usability and feasibility, where primary data were drawn from:

1. GEQ [18]: An adapted 32-item version of the basic module of the GEQ (skipping item 3 in the original 33-item) was used to measure seven aspects of game experience: flow, competence, positive affect, negative affect, tension/annoyance, challenge, and sensory/imaginative immersion. Item 3 ("I was interested in the game's story") was excluded from the component immersion because our own game did not have a story. Component flow and competence scores were of particular importance to intrinsic motivation. Component scores were calculated as the means of their items on a scale ranging from 0 to 4.
2. The Virtual Reality Sickness Questionnaire (VRSQ) [28]: It was used to detect and quantify physiological symptom indicators of distress that users may experience while controlling VR/MR hardware, such as eye strain, nausea, headache, disorientation, and dizziness. It is a 9-item measure scored on a 0- to 3-point scale, and the VRSQ total score was calculated as the mean of these items, which provided a final score between 0 and 3. It has been shown to be a valid measure in the original investigation of the VRSQ for cybersickness because it was very highly correlated with existing motion sickness indices.
3. Objective game performance logs: Primary objective metrics were derived from automatically collected gameplay data provided by the game's backend system. They were "game score," "movement accuracy," "mean time per round," "task completion rate," "movement level (extracted from hand movement and head movement trajectory)," and "cumulative game time." These data were direct evidence to substantiate the assessment of whether users could use the game in an efficient manner.

We took an integrated approach, triangulating subjective experience (GEQ), physical comfort (VRSQ), and actual performance (game logs) to provide a sound foundation for a unified and rigorous assessment of the game's usability and feasibility.

We adopted a target engineering level of 80% movement precision to be used as an internal and exploration benchmarking.

Psychometric Properties of Outcome Scales

Internal consistency was evaluated on the present sample using Cronbach α and Spearman-Brown split-half coefficients. As several components (tension/annoyance, challenge, and negative affect) exhibited pronounced floor effects with near-zero variance, internal consistency coefficients for these scales were not estimable; for these, we report distributional summaries only. Reliability results for all measures are presented in Table 2. Per the GEQ guidelines, we dropped the story-related immersion item (item 3), leaving 5 items for this component; reliability values in Table 2 reflect this adapted set.

Table . Internal consistency of Game Experience Questionnaire and Virtual Reality Sickness Questionnaire (VRSQ) subscales.

Measure	k (items)	Cronbach α	Split-half (Spearman-Brown)
Sensory and imaginative immersion	5[a]	0.657	0.583
Flow	5	0.546	0.539
Tension/annoyance	3	N/A[b] (floor effect)	N/A[b] (floor effect)
Challenge	5	N/A[b] (floor effect)	N/A[b] (floor effect)
Negative affect	4	N/A[b] (floor effect)	N/A[b] (floor effect)
Positive affect	5	0.86	0.85
VRSQ total	9	0.782	0.689

[a]Game Experience Questionnaire core scored 0 - 4; immersion used items 12, 18, 19, 27, 30 (item 3 removed). Split-half denotes Spearman-Brown.
[b]N/A indicates nonestimable due to floor effect (near-zero variance).

Operational Definitions

The key game performance indicators were defined as follows: score was the system-generated composite points; accuracy (%) was the proportion of frames where hotspots were within tolerance; and posture (count) represented the total number of successful hotspot holds. Per movement, the player aligns the left and right controllers and the HMD with 3 predefined hotspots; holding any hotspot for ≥3 seconds is counted as one successful hotspot hold. The 16-movement sequence loops whenever it is completed before the 5-minute limit; therefore, posture is an open-ended count (no fixed cap) within the time window. For interpretability, 48 holds correspond to one full loop (16 movements×3 hotspots).

Data Preparation and Statistical Analysis

Overview

Consistent with the a priori plan, no confirmatory between-condition hypothesis testing (main effects or interaction) was conducted. To enhance transparency, per-condition descriptive summaries (medians and IQRs for the 4 cells) are provided in Table 3.

Table . Per-condition descriptive statistics.

Outcome[a]	Group A: VR[b] soothing (n=17), median (IQR)	Group C: VR intense (n=19), median (IQR)	Group B: MR[c] soothing (n=17), median (IQR)	Group D: MR intense (n=17), median (IQR)
Positive affect	4 (3.8 - 4)	4 (3.8 - 4)	4 (3.4 - 4)	4 (3.8 - 4)
Flow	4 (3.2 - 4)	3.8 (3.6 - 4)	4 (3.2 - 4)	4 (3.2 - 4)
Competence	3.6 (3.1 - 4)	3.8 (3.2 - 4)	3.6 (3.4 - 4)	4 (3.3 - 4)
Challenge	1.4 (1.2 - 1.6)	1.4 (1 - 1.8)	1.6 (1.1 - 1.7)	1.6 (0.8 - 1.7)
VRSQ[d] total	0 (0 - 1)	0 (0 - 2)	0 (0 - 1.5)	0 (0 - 2)
Accuracy (%)	75 (56-84)	65 (54-81)	65 (54-77)	79 (54-94)

[a]Per-condition n shown in column headers. No inferential tests were planned; descriptive summaries only.
[b]VR: virtual reality.
[c]MR: mixed reality.
[d]VRSQ: Virtual Reality Sickness Questionnaire.

Data Cleaning

Handled sporadic missingness via median imputation only; no multiple imputation was performed [29].

Analysis Plan

Data normality was examined using the Shapiro-Wilk test [30,31]. Most key variables were found not to be normally distributed (Table 4). Nonparametric methods were consequently unanimously selected across hypothesis testing to yield robust and valid inferences. That is, median dimension scores on the GEQ were compared against a theoretical midpoint score of 2.0 using 1-sample Wilcoxon signed-rank tests. Wilcoxon tests correspondingly compared the median against a theoretical score of 0 on the VRSQ symptom scores (rated 0 - 3) to determine whether symptoms were distinctly present. With the exploratory nature of this item-level analysis, no adjustment for multiple comparisons was made on these 9 tests. Spearman ρ was used to examine associations between key factors. A level of significance was predetermined at an α of .05.

Table . Shapiro-Wilk test of normality for Game Experience Questionnaire (GEQ) dimensions, Virtual Reality Sickness Questionnaire (VRSQ) symptoms, and game performance indicators.

	W	N	*P* value
Overall GEQ dimension score			
Competence	0.798	70	<.001
Sensory and imaginative immersion	0.777	70	<.001
Flow	0.719	70	<.001
Tension/annoyance	0.275	70	<.001
Challenge	0.974	70	.16
Negative affect	0.573	70	<.001
Positive affect	0.551	70	<.001
VRSQ main symptom			
General discomfort	0.205	70	<.001
Fatigue	0.28	70	<.001
Headache	0.158	70	<.001
Eye fatigue	0.431	70	<.001
Difficulty concentrating	0.153	70	<.001
Head fullness	0.346	70	<.001
Blurred vision	0.349	70	<.001
Dizziness when closing eyes	0.512	70	<.001
Dizziness/vertigo	0.494	70	<.001
Key game performance			
Score	0.967	70	.06
Posture	0.967	70	.06
Accuracy	0.962	70	.03
Average time spent	0.796	70	<.001
Movement distance (m)	0.979	70	.29

Overall Usability Analysis

Descriptive statistics summarized the GEQ, the VRSQ, and game log indicators as median (IQR) [19,32].

Results

Participant Flow and Analysis Overview

In total, 86 older adults were recruited; 16 (18.6%) discontinued during acclimation due to intolerance, and data from 70 (81.4%) participants were analyzed. Participant flow and group allocation are shown in Figure 1. Per-condition medians (IQRs) for the 4 cells are available in Table 3.

Scale Reliability

Internal consistency was good for positive affect (α=.860; split-half=0.850) and acceptable for the VRSQ total (α=.782; split-half=0.689). Competence (α=.667; split-half=0.750) and sensory/imaginative immersion (α=.657; split-half=0.583) were in the moderate range, whereas flow was borderline (α=.546; split-half=0.539). Coefficients for tension/annoyance, challenge, and negative affect were not estimable due to floor effects (near-zero variance; Table 2).

Before the main analysis, the assumption of normality for the key-dependent variables was assessed using the Shapiro-Wilk test. The results are detailed in Table 4. While a few indicators did not significantly deviate from normality (eg, *challenge*, *P*=.16; game log *score*, *P*=.06; *posture*, *P*=.06; *movement distance (m)* *P*=.29), the majority of variables substantially deviated (eg, *accuracy*, *P*=.032; most GEQ/VRSQ dimensions, *P*<.001). Therefore, to maintain consistency and ensure robust inference across all analyses, nonparametric procedures were uniformly adopted.

Descriptive Statistics

We reported per-condition descriptive statistics for key outcomes across the 2×2 design (VR/MR×feedback intensity). As planned, no between-condition inferential tests were conducted; values are summarized as median (IQR) for interpretability with nonnormal distributions (see Table 3 for the 4 cells).

Overall Usability and Acceptability of the VR Tai Chi Game

To evaluate the overall usability and acceptability of the VR Tai Chi game for older users, this study collected and analyzed survey data from GEQ, VRSQ, and objective game performance log data.

Subjective Game Experience (GEQ)

As reflected in Table 5, subjective gaming experience as measured using the GEQ was highly positive. The median levels on positive aspects like "positive affect" (median 4.0), "flow" (median 4.0), "competence" (median 3.8), and "sensory/imaginative immersion" (median 3.8) all far surpassed the theoretical midpoint of the scale (*P*<.001 in all). Medians on negative aspects such as "tension/annoyance" (median 0) and "negative affect" (median 0) were below midpoint to a statistically significant degree (*P*<.001 in both). The median level on "challenge" was low to a statistically significant degree as well (median 1.4, *P*<.001).

Table . Descriptive statistics and one-sample Wilcoxon signed-rank test for Game Experience Questionnaire dimensions.

Dimension	Values, median (IQR[a])	*P* value[b]
Competence	3.8 (3.2 - 4)	<.001
Sensory/imaginative immersion	3.8 (3.2 - 4)	<.001
Flow	4 (3.2 - 4)	<.001
Tension/annoyance	0 (0 - 0)	<.001
Challenge	1.4 (1 - 1.6)	<.001
Negative affect	0 (0 - 0.25)	<.001
Positive affect	4 (3.8 - 4)	<.001

[a]Game Experience Questionnaire scores were rated on a 0 - 4 scale.

[b]*P* values reflect a one-sample Wilcoxon signed-rank test comparing the median score against the theoretical midpoint of 2.0.

Physiological Comfort (VRSQ)

The physiological comfort of the screened participants was assessed using the VRSQ. Overall, the incidence and severity of reported symptoms were low. Detailed item-level medians, statistical test results, incidence rates, and severity scores for reporters are presented in Table 6. The *P* values in the table should be interpreted as exploratory signals of which symptoms were most frequently endorsed, rather than as confirmatory hypothesis tests.

Table . Overall Virtual Reality Sickness Questionnaire main symptom statistical analysis.

Symptom item	Median	*P* value	Incidence rate (%)	Mean severity for reporters
General discomfort	0 (0-0)	.08	4.3	1
Fatigue	0 (0-0)	.04	7.1	1.4
Headache	0 (0-0)	.16	2.9	1
Eye fatigue	0 (0-0)	.002	15.7	1.273
Difficulty concentrating	0 (0-0)	.18	2.9	1.5
Head fullness	0 (0-0)	.01	10.0	1.286
Blurred vision	0 (0-0)	.01	10.0	1.143
Dizziness when closing eyes	0 (0-0)	<.001	20.0	1.214
Dizziness/vertigo	0 (0-0)	<.001	18.6	1

Game Performance

A look at the in-game objective performance data listed in Table 7 reveals wide individual variations among older adult participants. Participants demonstrated a fluid interaction rhythm with a median time spent per action of 6.4 seconds. However, large variations across major indicators were apparent due to wide IQRs seen across score and accuracy measurements (Table 7). Median movement accuracy was 68.5% (IQR 54.2% - 82.3%). Scores higher than 48 indicate that participants completed more than a single loop (or accrued numerous hotspot holds within a movement) within a 5-minute period.

Table . Key game performance indicators.

Indicator	Values, median (IQR)
Score	96 (70.25-133.5)
Posture[a] (count; no fixed cap, 48=one loop)	47 (39.75-55)
Accuracy (%)	68.5 (54.2 - 82.3)
Average time spent (s)	6.4 (5.5-7.5)
Movement distance (m)	273.5 (231.5-317.25)

[a]Posture counts the total hotspot holds within the retained 5-min session (best-of-two). The 16-movement sequence loops if completed early.

Correlation Analysis of Key Usability Factors

Correlation between major usability factors was examined using Spearman rank correlation analysis, and the results were presented in Table 8. Flow was significantly positively correlated with accuracy (ρ=.342; *P*=.004), and competence was significantly positively correlated with accuracy (ρ=.322; *P*=.007). VRSQ total score was significantly inversely correlated to positive affect (ρ=−0.334; *P*=.005). Competence was not significantly correlated with age (ρ=−0.179; *P*=.14).

Table . Spearman ρ correlation analysis (n=70).

Hypothesis	Variable 1	Variable 2	Correlation coefficient	*P* value
H1	Flow	Accuracy	0.342[a]	.004
H2	VRSQ[b] total score	Positive affect	−0.334[a]	.005
H3	Accuracy	Competence	0.322[a]	.007
H4	Age	Competence	−0.179	.14

[a]Correlation is significant at the 0.01 level (2 tailed).
[b]VRSQ: Virtual Reality Sickness Questionnaire.

Discussion

Subjective Experience Profile: High Perceived Competence With Low Perceived Challenge

As presented in Table 5, medians for positive affect, flow, and competence were above the scale midpoint, whereas challenge was markedly below (all *P*<.001). We observe a distinct user experience profile: high perceived competence and low challenge. For older adults who begin using exergames at home, this combination may be most relevant in reducing frustration and promoting first-time engagement [33,34]. At the same time, future research should gradually increase difficulty in an adaptive manner such that motivational support is balanced against accurate skill acquisition in free-living situations [35,36].

The internal consistency profile reveals adequate internal consistency on positive affect and satisfactory consistency on total VRSQ and moderate-to-borderline level in several components of the GEQ. Borderline α on flow is likely owing to constrained variance and the brevity and heterogeneity of item material in this first-stage application involving a single session; such findings on these components should consequently be interpreted correspondingly cautiously.

Physiological Comfort (VRSQ): Acceptability in Screened Sample and Potential Challenges

Regarding physiological comfort, the VR Tai Chi game demonstrated good tolerability within the screened participant sample, but this conclusion must be interpreted with caution [37]. The initial exclusion of a high 18.6% of recruits due to inability to adapt or strong VR-related sickness signifies a considerable bias in the VR tolerance of the final analyzed sample. Consequently, the currently observed low levels of cybersickness might overestimate the game's physiological comfort among the general older adult population. Although within the screened "more tolerant" sample, the average reported levels and perceived severity of various physiological discomfort symptoms (mainly mild "dizziness when closing eyes," "vertigo," and "eye fatigue") were low, 15% to 20% of participants still reported these mild symptoms [38]. This implies that cybersickness continues to be a primary barrier that has to be cleared ahead of VR technology deployment to older age categories. This 18.6% exclusion rate translates directly to current technological or design limitations in making VR universally comfortable among older adults and places a realistic limit on such interventions' immediate impact absent substantive improvements. For plans to promote such VR/MR Tai Chi games to the home environments of community-dwelling older adults, this high initial exclusion rate reveals key practical challenges. In home settings lacking immediate professional guidance, effective user screening, clear adaptation training guidelines, and easy-to-operate personalized comfort settings (eg, mode switching, visual effect adjustments) will be prerequisites for ensuring user safety and continued use [15,39]. To mitigate cybersickness at home, key strategies include starting with MR or low-vection scenes, using brief (≈5 min) sessions, ensuring proper headset fit with stable frame rates, and avoiding artificial locomotion. Furthermore, even mild symptoms, when experienced during unsupervised long-term home use, could affect user experience and adherence, thus necessitating continuous design optimization to minimize discomfort as much as possible [40].

Accuracy-Competence Paradox

Despite a median movement accuracy of 68.5% (IQR 54.2% - 82.3%), which was below our predefined exploratory benchmark of 80%, participants reported high perceived competence (median 3.8), reflecting a rule set that credits partial hotspot matches (≥1/3 for ≥3 s). This system focuses on older users' positive affective experiences to improve long-term exercise adherence while reducing frustration and elevating confidence but without concomitant increases in stricter movement norms. However, this dilemma positions an additional discussion regarding such a game's effectiveness as a very precise device for training within corresponding trade-offs between motivation and skill fidelity.

The permissive rule (≥1 hotspot for ≥3 s) retains motivation but tends to overcredit partial matches. Future iterations will include adaptive multipoint matching and more informative corrective cues to put verifiable and perceived competence into alignment while retaining engagement [41]. It presents a foundational tension: the probable discrepancy between perceived benefit/competence and verifiable physiological/skill-based benefit [42]. While perceived competence is central to adherence, if it is not adequately correlated to correct movement execution, long-term benefit to health can be undermined and signal a need for "verifiable competence" feedback loops that remain motivating. For a mass market product like this, selling independent home-based exercise to older people, such a motivational design can serve to buffer against users experiencing "anxiety alone" if having trouble and will invite continued participation [35]. It does raise a valuable question, though: without professional instruction in a home environment setup, how can it be ensured that movements/movement effectiveness is appropriate? Therefore, future iterations of the gameplay might involve implementing tiered success criteria or optional stepwise movement guidance and feedback modules such that users can reach a personally optimal balance between enjoying playing the game and checking movement correctness in accordance with their own needs [43].

Insights From Correlation Analysis of Key Usability Factors

The interplay between gameplay performance, user comfort levels, and subjective experience yields direct design recommendations for successful home-based exergames. It was found that "accuracy" was strongly positively correlated with "flow" and "competence." This indicates that more precise movement performance can guide users into a state of deep, immersive focus, forming a positive feedback loop of "successful execution → flow experience → enhanced participation." This also offers a more nuanced interpretation of the "accuracy-competence paradox": the game provides a high baseline of perceived competence for all participants through its lenient judgment, and on top of this, higher accuracy leads to an even stronger sense of competence, cleverly balancing universal motivation with individual rewards. This suggests that users do derive satisfaction from objective mastery; while a safety net of easy success is important, completely removing skill development or clear performance feedback might make the game less engaging long term for some [44,45]. Furthermore, the analysis also highlighted the impact of users' individual states: the degree of physiological discomfort (VRSQ total score) was significantly negatively correlated with "positive affect." This warns that even mild discomfort can negatively affect older users' enjoyment, which is particularly critical in home settings lacking immediate help, thereby re-emphasizing the necessity of minimizing cybersickness risks [46]. Finally, a positive signal was that the study found no significant relationship between "age" and "competence," suggesting that the game possesses good age inclusivity within the 60 to 75 years age group, and users' perceived competence did not systematically diminish with increasing age. This has positive implications for its promotion among diverse community-dwelling older adult groups, although caution is advised when generalizing this conclusion to very old adults [39,47].

Limitations

A number of limitations can be applied to this work. A selection bias in this sample exists since a large number of preliminary participants were lost due to side effects associated with VR, and the resulting group was predominantly female, younger, and physically active. This extremely high level of preexisting physical activity may not be characteristic among this typically sedentary population that such home-based exergames aim to reach but may not serve adequately and hence might limit how much our findings generalize to less active older individuals in this regard. This may inflate physiological comfort but restrict findings' generalizability to a wider range of heterogeneous older people, especially older individuals who may be less technology-savvy or physically constrained. Finally, a single, brief VR intervention cannot eliminate any novelty effects or long-term adherence or long-term effects on health and retention of skill in a naturalistic home environment.

Future Direction

Future research will need to proceed beyond virtual-simulated laboratory settings to home-based settings and use strategies such as Ecological Momentary Assessment to assess long-term use, adherence, skill learning, and actual resultant health effects. It will be important to diversify the samples of participants to include a broader age range, a wider range of health statuses, and a wider range of technology backgrounds to assess the generalizability of such findings. Studies will need to work toward developing adaptive VR systems capable of shifting difficulty and feedback in a dynamic fashion, dependent upon user progress. We will assess fixed versus adaptive thresholds and include per-user calibration routines within a home-based study.

This will aid in connecting, keeping participants motivated, and ensuring appropriate skill acquisition to promote successful unsupervised home-based practice. Although this current system used the Meta Quest 3 headset and hand tracking, future work could include alternative systems (eg, Xbox/Kinect) or outward-facing full-body tracking systems to better assess posture in unsupervised home-based settings. Ecological Momentary Assessment is described as a future longitudinal assessment methodology to assess adherence and in-the-wild engagement but not a single-session assessment of emotion.

Conclusions

This study demonstrates that well-designed VR/MR Tai Chi exercises can be a realistic and inspiring alternative exercise among community-dwelling older adults, supporting aging in place initiatives. A self-reported profile of high competence but low challenge perception implies early initiation but will need variable difficulty to maintain motivation in alignment with acuity of therapy for long-term use at home. Although physiologically tolerable in this prescreened population, a high entry exclusion rate due to cybersickness is a major barrier to wider use. A future direction to wider use success is a balance between motivational design of content and therapy acuity and modifications to system usability and comfort to provide net health gains in homeowner unsupervised use.

Acknowledgments

The authors are very grateful to the older residents of Chang’an Community in Changde City, Hunan Province, for participating in the experiment.

Funding

Funding was partly supported by a research grant from Universiti Kebangsaan Malaysia (GUP-2019-066).

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to participant privacy and confidentiality considerations and institutional ethics restrictions. Deidentified data are available from the corresponding author upon reasonable request for noncommercial research purposes, subject to approval by the relevant ethics committee where applicable and the signing of a data use agreement. The analysis code used in this study is available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: XS, NMA, MHMS, MYR
Methodology: XS
Formal analysis: XS
Investigation: XS
Data curation: XS
Writing – original draft preparation: XS
Writing – review & editing: all authors
Supervision: NMA, MHMS, MYR
Funding acquisition: NMA
Resources: NMA

Conflicts of Interest

None declared.

References

1. Au A, Lai S, Ng E. Introduction: mental health of older adults in Asia. Clin Gerontol 2015 May 27;38(3):187-189. [doi: 10.1080/07317115.2015.1008350]
2. Guastafierro E, Rocco I, Quintas R, et al. Identification of determinants of healthy ageing in Italy: results from the national survey IDAGIT. Ageing Soc 2022 Aug;42(8):1760-1780. [doi: 10.1017/S0144686X20001671]
3. Penger S, Grah B. Healthy aging and physical activity: the qualitative research. ENTRENOVA 2024;10(1):126-139. [doi: 10.54820/entrenova-2024-0012]
4. Mohd Johari NF, Mohamad Ali N, Mhd Salim MH, Abdullah NA. Factors driving the use of mobile health app: insights from a survey. mHealth 2025;11. [doi: 10.21037/mhealth-24-44] [Medline: 40248757]
5. Bajdek N, Latham NK, Dishaw M, et al. Feasibility of a multicomponent digital fall prevention exercise intervention for at-risk older adults. J Frailty Aging 2024;13(4):349-358. [doi: 10.14283/jfa.2024.35] [Medline: 39574254]
6. Huang X, Ali NM, Sahrani S. Evolution and future of serious game technology for older adults. Information 2024;15(7):385. [doi: 10.3390/info15070385]
7. Zhang W, Sun J, Feng X, Zhang H, Zhang Y, Zhao M. Effectiveness of Tai Chi exercise on fear of falling and balance in older adults: a meta-analysis. Geriatr Nurs 2023;51:194-201. [doi: 10.1016/j.gerinurse.2023.03.019] [Medline: 37011491]
8. Boot W, Dilanchian AT. Potential and pitfalls in the use of extended reality to support older adults. Innov Aging 2022;6(Suppl 1):216. [doi: 10.1093/geroni/igac059.860]

9. Wang PG, Ali NM, Sarker MR. A bibliometric analysis exploring the acceptance of virtual reality among older adults: a review. Computers 2024;13(10):262. [doi: 10.3390/computers13100262]
10. Chen M, Tang Q, Xu S, Leng P, Pan Z. Design and evaluation of an augmented reality-based exergame system to reduce fall risk in the elderly. Int J Environ Res Public Health 2020 Oct 1;17(19). [doi: 10.3390/ijerph17197208] [Medline: 33019759]
11. Huang X, Ali NM, Sahrani S. Haptic-driven serious card games for older adults: user preferences study. JMIR Serious Games 2025 May 27;13:e73135. [doi: 10.2196/73135] [Medline: 40424615]
12. Dyke G. The logical structure of a research paper in clinical healthcare. Biosis Biol Syst 2020;1(4):126-133. [doi: 10.37819/biosis.001.04.0088]
13. Song X, Ali NM, Mhd Salim MH, Rezaldi MY. A literature review of virtual reality exergames for older adults: enhancing physical, cognitive, and social health. Appl Sci (Basel) 2025;15(1):351. [doi: 10.3390/app15010351]
14. Benham S, Pennell K, Wynarczuk K. The meaning of immersive virtual reality from the perspectives of older adults: a qualitative analysis. Arch Phys Med Rehabil 2022 Mar;103(3):e17-e18. [doi: 10.1016/j.apmr.2022.01.047]
15. Brown JA, Dinh AT, Oh C. Safety and ethical considerations when designing a virtual reality study with older adult participants. 2022 Presented at: Proceedings of the 24th International Conference on Human–Computer Interaction (HCII 2022); Jun 26 to Jul 1, 2022. [doi: 10.1007/978-3-031-05581-2_2]
16. Hunicke R, LeBlanc M, Zubek R. MDA: a formal approach to game design and game research. Presented at: Proceedings of the AAAI-04 Workshop on Challenges in Game AI; Jul 25-26, 2004 URL: https://users.cs.northwestern.edu/~hunicke/MDA.pdf [accessed 2026-01-10]
17. Fu Y, Hu Y, Sundstedt V, Forsell Y. Conceptual design of an extended reality exercise game for the elderly. Appl Sci 2022;12(13):6436. [doi: 10.3390/app12136436]
18. IJsselsteijn WA, Kort YAW, Poels K. The game experience questionnaire. : Technische Universiteit Eindhoven; 2013 URL: https://pure.tue.nl/ws/portalfiles/portal/21666907/Game_Experience_Questionnaire_English.pdf?utm [accessed 2026-01-10]
19. Singh M, Sun D. Evaluating Minecraft as a game-based metaverse platform: exploring gaming experience, social presence, and STEM outcomes. Interact Learn Environ 2025;33(7):4272-4294. [doi: 10.1080/10494820.2025.2459200]
20. Fuentes J, Restrepo L, Reyes B, Sánchez R, Figueroa PA. Answering anya: a virtual and mixed reality game centered on nurturing. Presented at: Entertainment Computing - ICEC 2024 - 23rd IFIP TC 14 International Conference; Sep 30 to Oct 3, 2024. [doi: 10.1007/978-3-031-74353-5_28]
21. Delgado F, Greenberg J. Cognitive and practice effects of immersive virtual reality use in older adults: preliminary results. Innov Aging 2024 Dec 31;8(Suppl 1):1231. [doi: 10.1093/geroni/igae098.3939]
22. Ali NA, Hudson J, Stratton G, Mullins J. O.1.2-7 Optimization of older adults home spaces. Eur J Public Health 2023 Sep 11;33(suppl 1). [doi: 10.1093/eurpub/ckad133.094]
23. Shewaga R, Uribe-Quevedo A, Kapralos B, Alam F. A comparison of seated and room-scale virtual reality in a serious game for epidural preparation. IEEE Trans Emerg Topics Comput 2020;8(1):218-232. [doi: 10.1109/TETC.2017.2746085]
24. Chau PH, Kwok YYJ, Chan MKM, et al. Feasibility, acceptability, and efficacy of virtual reality training for older adults and people with disabilities: single-arm pre-post study. J Med Internet Res 2021 May 4;23(5):e27640. [doi: 10.2196/27640] [Medline: 33944795]
25. Xu TB, Mostafavi A, Boot WR, Czaja S, Kalantari S. Assessing the feasibility and efficacy of virtual reality navigational training for older adults. Innov Aging 2024;9(1):igae099. [doi: 10.1093/geroni/igae099] [Medline: 39781245]
26. Küntzer L, Schwab S, Spaderna H, Rock G. Measuring user experience of older adults during virtual reality exergaming. Presented at: 2024 16th International Conference on Quality of Multimedia Experience (QoMEX); Jun 18-20, 2024; Karlshamn, Sweden p. 153-159. [doi: 10.1109/QoMEX61742.2024.10598263]
27. Tian F, Ni S, Zhang X, et al. Enhancing Tai Chi training system: towards group-based and hyper-realistic training experiences. IEEE Trans Vis Comput Graph 2024 May;30(5):1-11. [doi: 10.1109/TVCG.2024.3372099] [Medline: 38457324]
28. Kim HK, Park J, Choi Y, Choe M. Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. Appl Ergon 2018 May;69:66-73. [doi: 10.1016/j.apergo.2017.12.016] [Medline: 29477332]
29. Borrohou S, Fissoune R, Badir H. Data cleaning survey and challenges – improving outlier detection algorithm in machine learning. J Smart Cities Soc 2023 Oct 9;2(3):125-140. [doi: 10.3233/SCS-230008]
30. Ahmed HS. Inferential statistics for cardiothoracic surgeons: part 3 - drawing valid conclusions from clinical data. Indian J Thorac Cardiovasc Surg 2025 Feb;41(2):233-247. [doi: 10.1007/s12055-024-01867-7] [Medline: 39822865]
31. Habibzadeh F. Data distribution: normal or abnormal? J Korean Med Sci 2024 Jan 22;39(3):e35. [doi: 10.3346/jkms.2024.39.e35] [Medline: 38258367]
32. Boletsis C. A user experience questionnaire for VR locomotion: formulation and preliminary evaluation. 2020 Presented at: Augmented Reality, Virtual Reality, and Computer Graphics, AVR 2020; Sep 7-9, 2020. [doi: 10.1007/978-3-030-58465-8_11]
33. Pacheco TBF, de Medeiros CSP, de Oliveira VHB, Vieira ER, de Cavalcanti FAC. Effectiveness of exergames for improving mobility and balance in older adults: a systematic review and meta-analysis. Syst Rev 2020 Jul 18;9(1). [doi: 10.1186/s13643-020-01421-7] [Medline: 32682439]

34. Wong AKC, Bayuo J, Su JJ, et al. Effectiveness of the support from community health workers and health care professionals on the sustained use of wearable monitoring devices among community-dwelling older adults: feasibility randomized controlled trial. J Med Internet Res 2024 Nov 18;26:e52435. [doi: 10.2196/52435] [Medline: 39556810]
35. Mehrabi S, Muñoz JE, Basharat A, et al. Immersive virtual reality exergames to promote the well-being of community-dwelling older adults: protocol for a mixed methods pilot study. JMIR Res Protoc 2022 Jun 13;11(6):e32955. [doi: 10.2196/32955] [Medline: 35700014]
36. Martins AC, Quatorze J, Guia D. Exergaming as a tool to enhance strength, balance, gait, mobility, participation, self-efficacy for exercise and adherence in older adults. Eur J Public Health 2020 Jun 1;30(Suppl 2):024. [doi: 10.1093/eurpub/ckaa040.024]
37. Kim J, Kim Y, Chang PS, Min Oh S, Han S. A pilot study of virtual reality (VR) Tai Chi program on mental health among older adults during the COVID-19 pandemic. Am J Health Behav 2022 Oct 17;46(5):576-585. [doi: 10.5993/AJHB.46.5.8] [Medline: 36333829]
38. Pau M, Arippa F, Leban B, et al. Cybersickness in people with multiple sclerosis exposed to immersive virtual reality. Bioengineering (Basel) 2024 Jan 24;11(2):115. [doi: 10.3390/bioengineering11020115] [Medline: 38391601]
39. Mehrabi S, Basharat A, Mazen S, et al. Age in place, move in VR: investigating the feasibility and usability of a custom-made virtual reality exergame to promote the well-being of community-dwelling older adults. Alzheimer's & Dementia 2023 Dec;19(S23):e071610 [FREE Full text] [doi: 10.1002/alz.071610]
40. Rmadi H, Maillot P, Artico R, et al. Tolerance of immersive head-mounted virtual reality among older nursing home residents. Front Public Health 2023;11:1163484. [doi: 10.3389/fpubh.2023.1163484] [Medline: 37538272]
41. Ong DSM, Weibin MZ, Vallabhajosyula R. Serious games as rehabilitation tools in neurological conditions: a comprehensive review. Technol Health Care 2021;29(1):15-31. [doi: 10.3233/THC-202333] [Medline: 32804107]
42. Tondorf DF, Hounsell M, Pereira VA. Evaluating a serious game for math: is it useful and fun? JIS 2025;16(1):199-212. [doi: 10.5753/jis.2025.4656]
43. Guimarães V, Oliveira E, Carvalho A, et al. An exergame solution for personalized multicomponent training in older adults. Appl Sci 2021;11(17):7986. [doi: 10.3390/app11177986]
44. Guzmán DE, Rengifo CF, Guzmán JD, Garcia Cena CE. Virtual reality games for cognitive rehabilitation of older adults: a review of adaptive games, domains and techniques. Virtual Real 2024;28(2). [doi: 10.1007/s10055-024-00968-3]
45. Borrego G, Morán AL, Meza V, Orihuela-Espina F, Sucar LE. Key factors that influence the UX of a dual-player game for the cognitive stimulation and motor rehabilitation of older adults. Univ Access Inf Soc 2021 Nov;20(4):767-783. [doi: 10.1007/s10209-020-00746-3]
46. Vlahovic S, Skorin-Kapov L, Suznjevic M, Pavlin-Bernardic N. Not just cybersickness: short-term effects of popular VR game mechanics on physical discomfort and reaction time. Virtual Real 2024;28(2):1-30. [doi: 10.1007/s10055-024-01007-x]
47. Wang YH. Understanding senior adults' needs, preferences, and experiences of commercial exergames for health: Usability study. JMIR Serious Games 2024 Apr 5;12:e36154. [doi: 10.2196/36154] [Medline: 38578674]

Abbreviations

GEQ: Game Experience Questionnaire
HMD: head-mounted display
MDA: mechanics, dynamics, and aesthetics
MR: mixed reality
VR: virtual reality
VRSQ: Virtual Reality Sickness Questionnaire

Edited by N Ahmadpour; submitted 23.Jun.2025; peer-reviewed by JAA Bakar, T Baranowski, X Liang; revised version received 08.Sep.2025; accepted 30.Oct.2025; published 28.Jan.2026.

Please cite as:
Song X, Ali NM, Salim MHM, Rezaldi MY
Immersive Tai Chi for Home-Based Exercise in Older Adults: Usability and Feasibility Study
JMIR Serious Games 2026;14:e79453
URL: https://games.jmir.org/2026/1/e79453
doi:10.2196/79453

Original Paper

Effects of Game-Based Learning on Piano Music Knowledge Among Elementary School Pupils: Pretest-Posttest Quasi-Experimental Study

Yun Wang[1,2*], BA, MA, PhD; Wee Hoe Tan[1,3*], BMm, MMM, PhD; Qiaoling Ye[4*], BA; Tingting Gu[5], BA, MA

[1]Faculty of Social Sciences and Liberal Arts, UCSI University, Cheras, Kuala Lumpur, Malaysia

[2]College of Music and Education, Chuzhou University, Chuzhou, Anhui, China

[3]Chongqing University of Posts and Telecommunications, Chongqing, Chongqing, China

[4]Chengbei Primary School, Chaohu, Anhui, China

[5]Faculty of Teacher Education, Chaohu University, Chaohu, Anhui, China

[*]these authors contributed equally

Corresponding Author:
Wee Hoe Tan, BMm, MMM, PhD
Faculty of Social Sciences and Liberal Arts
UCSI University
1, Jalan UCSI, UCSI Heights (Taman Connaught)
Cheras, Kuala Lumpur, 56000
Malaysia
Phone: 60 391018880
Fax: 60 391023606
Email: Tanwh@ucsiuniversity.edu.my

Abstract

Background: Music education is central to holistic child development; yet, traditional piano instruction in China often emphasizes rote memorization at the expense of meaningful understanding. Elementary school pupils, particularly novices, frequently struggle with rhythm, melody, and music theory. Game-based learning (GBL), which applies game elements to teaching, may address these challenges by enhancing engagement, providing immediate feedback, and fostering long-term learning.

Objective: This study aimed to examine the effects of a GBL module for piano education on elementary school pupils' music knowledge in Anhui Province, China, compared to traditional instruction.

Methods: A quasi-experimental, nonequivalent control group pretest-posttest design was used. Participants were allocated nonrandomly to experimental and control groups based on scheduling feasibility and teacher availability. A total of 60 novice piano learners (mean age 8.8 years, SD 1.16 years; 16 boys and 44 girls) from 3 elementary schools were assigned to either the GBL group (n=30) or the control group (n=30). Music knowledge was measured using the standardized Level 1 Basic Music Written Test before and after an 8-week intervention. Nonparametric analyses were applied, including Mann-Whitney *U*, Wilcoxon signed-rank, and McNemar item-level analyses. Rank-based effect sizes (*r*) and 95% CIs were reported where applicable.

Results: Baseline differences were present, with the control group scoring higher at pretest (median 52, IQR 24-76) than the GBL group (median 28, IQR 16-64; Mann-Whitney *U*=265.50; *r*=–0.35; 95% CI –0.39 to –0.32; *P*=.006). After the intervention, the GBL group significantly outperformed controls (median 100, IQR 88-100 vs median 60, IQR 40-92; Mann-Whitney *U*=4.0; *r*=–0.87; 95% CI –0.90 to –0.83; *P*<.001). Within-group analyses confirmed significant pre-post improvements for both groups (control *Z*=–3.24; *r*=–0.59; *P*=.001; and GBL *Z*=–4.81; *r*=–0.88; *P*<.001). Item-level McNemar tests showed significant improvements (*P*<.05) in 5 of 25 items for the GBL group. Missing data were negligible (<2%) and handled via listwise deletion after Little's missing completely at random (MCAR) confirmation (*P*=.08).

Conclusions: The GBL module significantly improved pupils' music knowledge, overcoming baseline disparities and producing posttest score gains with consistent mastery. The innovation of the study lies in the systematic integration of gamification with Orff and Dalcroze pedagogy through the Sidek Module Development Model, which distinguishes it from previous music education studies that examine gamification in isolation. By providing a validated, cost-effective, and scalable instructional module, the study contributes empirical evidence to the field of game-based music education and other practical implications for improving piano instruction in resource-constrained elementary school settings.

(JMIR Serious Games 2026;14:e80766) doi:10.2196/80766

KEYWORDS

quality education; gamified learning; piano education; music knowledge; school pupils

Introduction

Music education is a cornerstone of holistic development for elementary school pupils in China, nurturing not only artistic literacy but also cognitive, emotional, and social growth. The piano is highly regarded among musical instruments due to its versatility and extensive skill set. In recent decades, young pupils have taken up piano, a phenomenon known as the "piano heat" [1]. This increase shows a societal commitment to arts education, as piano skills enhance a child's cultural capital and allow for creative self-expression. The setting of this study, Anhui Province, has seen piano education grow along with national trends and demonstrate regional commitment to early artistic development. Despite the excitement for piano learning, it remains difficult to ensure that pupils develop a profound, meaningful understanding of music beyond mechanical proficiency.

Traditional piano teaching methods that emphasize rote memorization and mechanical practice typically fail to interest and comprehend young learners [2,3]. Due to their cognitive and emotional growth, elementary school pupils require pedagogical approaches that match their capacities. Their short attention spans and abstract nature make them struggle to grasp rhythm, melody, and basic music theory [4-6]. The lack of musical knowledge might also prevent pupils from connecting emotionally with the subject, reducing their motivation and interest [7]. These obstacles highlight the need for innovative teaching strategies that improve music knowledge and engagement.

Contemporary research identifies critical shortcomings in traditional piano beginner teaching, including a reliance on single-minded evaluation models, a neglect of process-based instruction, and a teaching focus that often prioritizes technical results over genuine artistic development. A key challenge remains the comprehensive cultivation of musical literacy, which extends far beyond the accurate reproduction of notes and rhythms [8]. To address these pedagogical gaps, game-based learning (GBL) has emerged as a promising intervention. GBL represents a fusion of informal and formal education, leveraging the interactive nature of games to deliver structured content. Scholarly reviews confirm that interactive mechanics of GBL, such as the use of challenges, rewards, and progress tracking, provide learners with a sense of autonomy and competence [9]. Furthermore, well-designed educational games can deepen a pupil's understanding of abstract content, encouraging them to find diverse solutions and train their creative and critical thinking skills. These mechanisms align directly with the goal of developing comprehensive musical literacy that transcends simple rote technique [10].

Gamified learning, which uses game design to boost engagement and learning, is a transformative solution. Reimagining piano exercises as engaging, entertaining experiences makes abstract concepts more approachable and fosters continuous involvement in music instruction [11]. Gamified learning works well in music education with obstacles, rapid feedback, and incentive systems [12]. Gamified learning also fits well with music pedagogy like the Orff and Dalcroze Eurhythmics, which emphasize active participation, rhythm, and movement to meet young learners' developmental needs [13-16].

In Anhui, elementary school pupils' piano music knowledge is still lacking despite research showing the benefits of gamified learning in other educational settings and music contexts. Regional specificity is essential for meaningful research since local teaching practices and resource availability may affect gamified intervention efficacy [17]. Gamified music education literature emphasizes short-term outcomes, like skill acquisition, but lacks data on long-term knowledge retention and conceptual understanding [10]. Few studies have examined how gamified learning can be integrated with traditional piano teaching methods or tailored to individual learner differences.

This study examines how a tailored GBL module affects elementary school pupils' music knowledge in Anhui. Innovative, evidence-based strategies to meet these learners' unique issues are used to promote music instruction. The design of the GBL module was grounded in established pedagogical principles of Orff-Schulwerk and Dalcroze Eurhythmics, which provided the theoretical foundation for integrating movement, rhythm, and embodied cognition into game tasks. These theories collectively informed the development of a multimodal, developmentally appropriate, and musically grounded GBL intervention for elementary pupils.

The findings should make theoretical and practical contributions to the field. The study expanded on Orff and Dalcroze frameworks to better understand how GBL can promote cognitive, emotional, and creative development in piano teaching. It will provide a validated GBL module for educators to enhance teaching, particularly in resource-constrained or outdated curriculum situations. The findings may also inform policymakers about how gamified learning might improve music education and offer a scalable approach. Accordingly, this study aims to examine the effects of a GBL module on piano music knowledge among elementary school pupils in Anhui province, China, and hypothesizes that pupils receiving the GBL intervention demonstrate greater improvements than those receiving traditional instruction.

Methods

Study Design

This study used a quasi-experimental, nonequivalent control group pretest-posttest design to examine the effects of a GBL module for piano education on elementary school pupils' music knowledge in Anhui Province, China. This design allowed comparisons between an intervention group receiving gamified piano instruction and a control group receiving traditional instruction. The report follows the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) statement to enhance transparency and reproducibility [18].

Setting

The study was conducted at a private piano training center located in Anhui Province, China. Three elementary schools partnered with the training center to provide access to novice piano learners. Data collection took place between February and April 2025, with 8 consecutive weekly sessions delivered during this period. Teacher workshops to standardize instructional methods were conducted in January 2025 before the intervention.

Participants

A total of 60 elementary school pupils aged 7 to 12 years were recruited through a series of briefing sessions held at the training center and affiliated schools. Eligibility criteria included (1) no previous formal piano training, (2) enrollment in the first year of piano lessons at the training center, and (3) parental consent and child assent for participation. Pupils with known hearing impairments or learning disabilities that could substantially affect music learning were excluded. Participants were systematically assigned to an experimental or a control group using a quasi-random sampling procedure: the first 60 pupils whose parents or guardians signed consent forms were enrolled, and every alternate consenting child was assigned to one of the two groups. This resulted in 30 participants in the gamified learning group and 30 in the control group. The final sample comprised 16 boys and 44 girls, consistent with broader trends in higher female participation in music education. Table 1 summarizes demographic characteristics. Although participants were alternately allocated, the procedure did not constitute true randomization. To minimize allocation bias, baseline demographics and pretest scores were compared across groups to assess equivalence.

Table 1. Demographic characteristics of elementary school pupils participating in a quasi-experimental study of game-based piano learning in Anhui, China.

Demographic variable	Experimental group, n	Control group, n	Total, n
Age (years)			60
7	2	4	6
8	6	17	23
9	7	6	13
10	11	3	14
11	3	0	3
12	1	0	1
Sex			60
Male	11	5	16
Female	19	25	44
Year in school			60
Year 1	3	9	12
Year 2	3	12	15
Year 3	13	9	22
Year 4	7	0	7
Year 5	3	0	3
Year 6	1	0	1

Intervention

The experimental group received an 8-week GBL module for piano education designed using the Sidek Module Development Model (SMDM). The module consisted of eight 45-minute sessions incorporating music games, including the "Central C Knocking Game," "Rhythm Clapping Game," and "Music Notation Song Game." Content followed a progressive trajectory from basic note recognition to more complex rhythm and score interpretation, guided by Orff and Dalcroze pedagogical principles. All lessons were delivered on an individual basis, with hands-on keyboard practice embedded within structured teacher-led activities. Attendance was recorded weekly, and teachers completed session-by-session fidelity checklists to ensure consistent adherence to the intended instructional sequence for both the gamified and traditional teaching conditions.

The control group received traditional piano instruction over the same period. Both groups were taught by the same 5 piano teachers, who attended a 3-day workshop to ensure standardized delivery of both instructional methods. Teachers applied gamified methods for the intervention group and conventional techniques for the control group. Neither participants nor teachers were blinded due to the nature of the educational intervention; however, outcome assessors scoring the written test were blinded to group allocation.

GBL Module Development

Overview

The SMDM, proposed by Sidek and Jamaludin [19], emphasizes systemic structure and the operational feasibility of instructional activities, making it particularly suitable for developing educational modules. In the context of a GBL module for piano education for elementary school pupils, the SMDM iterative and user-centered approach aligns seamlessly with GBL principles, fostering an engaging and effective learning experience tailored to young learners' needs. The structured yet flexible model of SMDM supports the creation of a GBL module for piano education by prioritizing user engagement and iterative refinement. The SMDM was used to create the GBL module to enhance elementary school pupils' piano music knowledge in Anhui [20]. The SMDM is a user-centered instructional design framework that emphasizes comprehensive needs analysis and iterative feedback to create effective multimedia learning tools. The model prioritizes user involvement and adaptability, which are essential for addressing the dynamic challenges in piano education, such as motivation, posture, and skill progression. While the original SMDM framework consists of 2 main stages, the study adapted it into 5 phases (design, development, assembly, validation, and evaluation) to provide a more granular and systematic approach tailored to the development of a GBL module for piano education. This expansion was necessary to accommodate the complexities of integrating technology, gamification, and pedagogical scaffolding in piano learning for elementary school pupils. The needs analysis stage was refined into the design phase for thorough user requirement gathering, while the multimedia development stage was subdivided into development, assembly, and validation phases to ensure iterative refinement and quality assurance. The evaluation phase was added as a distinct step to assess the effectiveness of the module after its implementation, aligning with empirical validation of SMDM. This 5-phase adaptation mirrors the scaffolding process in one-to-one piano lessons, where instruction progresses systematically from foundational to advanced skills, and supports targeted interventions. This systematic approach made the module pedagogically robust, engaging, and customized to young learners. Design, development, assembly, validation, and evaluation built on each other to create a cohesive educational tool.

Design Phase

The main goal of the module was to develop a dynamic GBL approach that elevates piano music knowledge for 7- to 12-year-olds. The aim was to craft a challenging but fun learning experience that deepened musical understanding. A game-based piano learning experience for elementary school pupils should be engaging and appealing [21]. The module emphasizes engaging experiences, clear goals, immediate feedback, diverse gaming techniques, and adaptive difficulty levels. Games that reinforce piano music knowledge use these principles. A comprehensive needs analysis identified learning challenges and ensured the module addressed them through targeted gamification. The 4Keys2Fun framework [22] was used to keep pupils engaged by incorporating social interaction, challenge, exploration, and purposeful learning to improve musical knowledge retention.

Development Phase

Content and instructional strategies were carefully selected during development to support a progressive learning trajectory. Music was used to teach specific knowledge outcomes, starting with note recognition and rhythm in "Please Play" and advancing to complex structures in "Church Organ" and "Yankee Doodle." Teacher demonstration, GBL, guided practice, rhythmic chanting, and formative feedback were used to teach piano. "Rhythm Clapping Game" (pupils competed to replicate rhythmic patterns in pairs, turning practice into a lively and collaborative challenge that reinforced knowledge while fostering peer learning), "Musical Notation Song Game," and "Central C Knocking Game" were embedded to improve rhythm, note positions, and key identification. A piano and copies of John Thomson's Easiest Piano Course 1 were needed for 45-minute sessions, ensuring the module is accessible. The teacher demonstrated real-time modeling to explain music.

Assembly and Draft Completion

A backward design model prioritized piano music knowledge outcomes before aligning assessments, content, and activities [23]. This concentrated and unified the framework. A total of 8 successive lessons with objectives, assessment tools, and resources guided pupils from basic note recognition to sophisticated musical interpretation in the final edition. This systematic assembly ensured that every part improved musical comprehension.

Validation and Pilot Testing

Validation was essential for module reliability and suitability. Three music pedagogy specialists critically assessed its content and design, followed by semistructured interviews with 2 piano teachers. Their suggestions improved content alignment and game dynamics. After expert approval, a pilot test with 17 students over 8 weeks (November 19, 2024, to January 14, 2025) provided practical input and revisions to ensure the readiness of the module for broader use. The 17 participants were novice learners aged 7 to 11 (mean age 8.8, SD 1.16) years, with 11 girls and 6 boys, mostly in Year 2 and Year 3 of elementary school.

Piano Education GBL Module

This basic school package has 8 lessons that establish musical principles through individual works. It uses games like "Central C Knocking Game," "Five-Line Speech Practice," and "Right-Hand Shooting Game" (students aimed to strike correct notes under time pressure, and their excitement translated into improved recognition and accuracy) to enhance learning. The curriculum systematically teaches piano key identification, note values, rests, and specific notes (C, D, B, E, A, G, and F). It emphasizes practical exercises to improve finger independence, rhythmic precision, and technical fluency on the piano. It also promotes curiosity, self-confidence, and music appreciation through interactive games and collaborative activities. Rhythm accuracy, hand coordination, and note and rest recognition are used to assess student progress, ensuring evidence-based musical progression. This structured technique improves musical proficiency and makes learning fun, supporting game-based music education.

Assessments and Outcomes

The primary outcome was music knowledge, assessed using the Level 1 Basic Music Written Test developed by the Chinese Music Academy. This 25-item multiple-choice test evaluates aural recognition (tones, melodic contours, dynamics, tempo, and rhythmic structures) and theoretical knowledge (pitch, solfège, intervals, meter, and musical symbols). The test was administered at baseline (pretest) and at the end of the intervention (posttest). Exposures included the gamified learning intervention, while potential confounders such as age, grade level, and baseline knowledge were recorded.

Study Size

The study included 60 participants (30 per group). The sample size was determined pragmatically based on the capacity of the training center and the number of eligible consenting pupils during the recruitment window. Although no formal power calculation was conducted, a sample of 30 per group has been shown in previous quasi-experimental studies to detect medium-to-large effects in educational outcomes.

Recruitment Procedures

Recruitment was conducted in collaboration with school administrators and the piano training center. Parents and guardians were provided with written study information, and informed consent was obtained prior to enrollment. Children also provided verbal assent. Recruitment was based on voluntary self-selection following briefing sessions conducted at the partner schools and training center. All eligible pupils whose parents provided consent during the 2-week recruitment window were included, and no incentives were offered.

A total of 86 pupils were screened for eligibility; 60 met the inclusion criteria and consented to participate. All 60 pupils were allocated to one of the two study conditions (30 experimental and 30 control). All participants received the assigned intervention, completed the 8 sessions, and completed both pretest and posttest assessments. No participants were lost to follow-up, and no protocol deviations occurred.

Data Analysis

All analyses were conducted using SPSS (version 29.0; IBM Corp). The normality of the pretest was assessed using the Shapiro-Wilk test, which indicated nonnormal distributions; hence, nonparametric statistical tests were chosen. Mann-Whitney *U* tests were carried out to compare the control and experimental groups in every pretest and posttest. Rank-based effect sizes (*r*) and 95% CIs were measured where applicable. Wilcoxon signed-rank tests examined changes from pretest to posttest within each group. McNemar tests assessed changes in correct responses for individual items on the 25-item test [24]. Data completeness was assessed, in which only less than 2% of responses were missing, and Little's missing completely at random (MCAR) test indicated data were missing completely at random (P=.078). As the proportion was small, listwise deletion was applied. No imputations were performed. Interpretations emphasized both statistical significance and precision of effect estimates, with CI reporting throughout.

Ethical Considerations

This study was approved by the Institutional Ethics Committee, UCSI University, under approval number IEC-2024-FOSSLA-0167, dated November 19, 2024. Written informed consent was obtained from all legal guardians, and verbal assent was provided by children before participation.

Confidentiality was ensured by anonymizing all datasets and storing them securely with restricted access. No personal identifiers were included in analyses or publications. Participants were not given financial incentives, but small educational gifts (eg, notebooks and stationery) were provided as tokens of appreciation. No identifiable photographs were taken; all supplementary materials were free of personal identifiers. Where identifiable images were unavoidable, written consent was obtained from both pupils and guardians.

Results

The pretest results showed a baseline piano music knowledge difference between groups (Table 2). The control and experimental groups had considerable piano music knowledge gaps before the intervention. The control group recorded a median pretest score of 52 (range 24-76) and a mean of 47.73 (SD 14.59), while the experimental group had 28 (range 16-64) and 36.13 (SD 12.92). In Table 3, a Mann-Whitney *U* test revealed a statistically significant difference (Mann-Whitney U=265.50; P=.006), indicating that the control group began with a higher level of knowledge.

Table 2. Pretest and posttest piano music knowledge scores among elementary school pupils receiving an 8-week game-based learning module versus traditional piano instruction in Anhui, China.

	Pretest scores			Posttest scores		
	Control[a]	Experimental[b]	Both	Control[a]	Experimental[b]	Both
n	30	30	60	30	30	60
Median (IQR)	52 (32-60)	32 (28-45)	40 (28-56)[c]	60 (51-65)	100 (95-100)	90 (60-100)[d]
Mode	60	28	28	60	100	100
Mean (SD)	47.73 (14.59)	36.13 (12.92)	41.93 (14.86)	58.67 (11.71)	97.33 (3.68)	78 (21.31)
Minimum	24	16	16	40	88	40
Maximum	76	64	76	92	100	100

[a]Control group (n=30).
[b]Experimental group (n=30).
[c]*P*=.006.
[d]*P*<.001.

Table 3. Between-group comparison of piano music knowledge at pretest and posttest using Mann-Whitney U tests in a quasi-experimental study of elementary school pupils in Anhui, China.

	Mann-Whitney *U*	*Z*	*r* (95% CI)	*P* value
Pretest	265.50	–2.75	0.35 (–0.39 to –0.32)	.006
Posttest	4.00	–6.71	0.87 (–0.90 to –0.83)	<.001

To evaluate the effectiveness of an 8-week GBL intervention on piano music knowledge, Mann-Whitney *U* tests were conducted to compare pretest and posttest scores between the experimental group (n=30) and the control group (n=30). Descriptive statistics and test results are presented in Table 3. No adverse events or unintended effects were observed in either study condition throughout the 8-week intervention.

In the pretest, a significant difference was observed between the experimental and control groups (Mann-Whitney *U*=265.50; *Z*=–2.75; *P*=.006), with the experimental group demonstrating higher scores compared to the control group. The effect size was medium (*r*=0.35, 95% CI –0.39 to –0.32), indicating a moderate baseline advantage for the experimental group.

In the posttest, the experimental group significantly outperformed the control group (Mann-Whitney *U*=4.00; *Z*=–6.71; *P*<.001). The experimental group achieved a median score of 100 (range 88-100; mean 97.33, SD 3.69), while the control group had a median score of 60 (range 40-92; mean 58.67, SD 11.71). The effect size was large (*r*=0.87; 95% CI –0.90 to –0.83), reflecting a substantial improvement in the performance of the experimental group. These results suggest that the 8-week GBL module effectively reversed the baseline performance patterns, closing the initial knowledge gap and significantly enhancing piano music knowledge in the experimental group compared to the control group.

The posttest scores achieved by the experimental group indicate a high degree of knowledge mastery, with reduced variability, in contrast to the more varied outcomes observed in the control group. The decreased SD (SD 3.69) in the experimental group suggests more consistent learning gains, reinforcing the ability of the GBL module to achieve uniform learning outcomes. Table 4 shows pretest and posttest score disparities demonstrating the effect of the intervention. The experimental group experienced a median gain of 64 points (range 36-84) with a mean improvement of 61.20 (SD 12.61). The pronounced disparity in knowledge gains shows that the GBL module promotes significant learning progress.

Table 4. Differences in piano music knowledge scores between pretest and posttest among elementary school pupils receiving game-based versus traditional piano instruction in Anhui, China

	Score difference between pretest and posttest		
	Control group	Experimental group	Control group and experimental group
n	30	30	60
Median (IQR)	8 (4-16)	64 (55-68)	36 (8-64)
Mode	4	64	64
Mean (SD)	10.93 (11.64)	61.20 (12.61)	36.07 (28.05)
Minimum	–4	36	–4
Maximum	64	84	84

To assess the effect of an 8-week GBL intervention on piano music knowledge, the Wilcoxon Signed Ranks test was conducted to compare pretest and posttest scores within the experimental group (n=30) and the control group (n=30). Results are presented in Table 5.

Table 5. Within-group comparison of pretest and posttest piano music knowledge using Wilcoxon signed-rank tests among elementary school pupils in Anhui, China.

Group	*Z*	*r* (95% CI)	*P* value
Control group	–3.24	–0.59 (–0.66 to –0.53)	.001
Experimental group	–4.81	–0.88 (–0.94 to –0.82)	<.001

Both groups demonstrated significant improvements in piano music knowledge from pretest to posttest. For the control group (*Z*=–3.24; *P*=.001), a substantial increase was observed (Z=–3.24; *P*=.001), with a large effect size (*r*=–0.59, 95% CI –0.66 to –0.53). The experimental group exhibited a more robust improvement (*Z*=–4.81; *P*<.001), with a larger effect size (*r*=–0.88, 95% CI –0.94 to –0.81). The higher *Z* value and lower *P* value in the experimental group with a larger effect size indicate that the GBL intervention led to greater knowledge growth compared to the control group. These findings, combined with the between-group comparison (see Table 3), suggest that the GBL module significantly enhanced learning outcomes, effectively closing the baseline knowledge gap in the experimental group.

McNemar tests were conducted on pretest and posttest scores for 25 test items to assess significant changes in performance. The item-level McNemar tests were exploratory ancillary analyses intended to examine differential item difficulty and the distributional impact of the intervention across 25 outcomes.

For the experimental group, McNemar tests were applied to each of the 25 piano knowledge items (Table 6). The analysis revealed statistically significant improvements (*P*<.05) in 5 items, highlighting the targeted effectiveness of the GBL module [25]. The remaining 20 items showed no statistically significant changes (*P*>.05).

Table 6. Item-level McNemar analyses of changes in piano music knowledge test items following an 8-week game-based learning (GBL) intervention among elementary school pupils in Anhui, China.

	Test items	Counts				Mode	*P* value
				Post[a]			
				W[b]	C[c]		
1	Listen and discern which sound is higher.					W-W	.04
		Pre[d]	W	13	3		
		—[e]	C	12	2		
2	Listen and discern which sound has a shorter duration.					W-W	≥.99
		Pre	W	11	6		
		—	C	7	6		
3	Listen and discern, choose the correct pitch.					W-W	.11
		Pre	W	16	2		
		—	C	8	4		
4	Listen and discern which one is the Concord interval.					W-W	.008
		Pre	W	21	0		
		—	C	8	1		
5	Comparing melody intervals, which interval is farther away?					C-W	.001
		Pre	W	10	2		
		—	C	16	2		
6	Listen and discern the direction of melody.					W-W	.27
		Pre	W	10	4		
		—	C	9	7		
7	Listen and discern the melody, choose the correct beat number.					C-C	.79
		Pre	W	5	6		
		—	C	8	11		
8	Listen and discern, choose the rhythm within the square.					W-W	N/A[f]
		Pre	W	16	0		
		—	C	14	0		
9	Listen and discern the melody, choose the correct note within the square.					W-W	.73
		Pre	W	18	3		
		—	C	5	4		
10	Listen to the melody. What is the style of this piece of music?					W-W; W-C	.80
		Pre	W	9	9		
		—	C	7	5		
11	Which note has a lower vocal range?					W-W	.79
		Pre	W	11	8		
		—	C	6	5		
12	Which one is a harmony interval?					W-W	.27

	Test items	Counts				Mode	*P* value
				Post[a]			
				W[b]	C[c]		
		Pre	W	13	4		
		—	C	9	4		
13	Choose the correct singing name.					W-C; C-C	N/A
		Pre	W	0	15		
		—	C	0	15		
14	Choose the correct interval degree.					C-W; C-C	.09
		Pre	W	7	3		
		—	C	10	10		
15	Which one is the dissonant interval?					C-C	.23
		Pre	W	9	3		
		—	C	8	10		
16	Which one is the sixteenth rest?					W-C; C-C	N/A
		Pre	W	0	15		
		—	C	0	15		
17	Choose the correct symbol meaning.					W-C	N/A
		Pre	W	0	16		
		—	C	0	14		
18	According to the spectrum example, choose the correct beat number.					W-C	<.001
		Pre	W	1	21		
		—	C	1	7		
19	According to the beat sign, how many bars of the score are shared?					W-C	.01
		Pre	W	5	12		
		—	C	2	11		
20	Choose the correct sound name and complete the C natural major scale.					C-C	.11
		Pre	W	3	8		
		—	C	2	17		
21	The instrument in the picture is _______.					C-C	N/A
		Pre	W	0	12		
		—	C	0	18		
22	The instrument in the picture is _______.					C-C	N/A
		Pre	W	0	12		
		—	C	0	18		
23	The instruments belonging to the woodwind group are_______.					W-C; C-C	N/A
		Pre	W	0	15		
		—	C	0	15		
24	Representatives of the “classical music” period.					W-C	N/A
		Pre	W	0	22		
		—	C	0	8		

	Test items	Counts				Mode	P value
				Post[a]			
				W[b]	C[c]		
25	"Sunset Flute and Drum" belongs to _______.					C-C	NA
		Pre	W	0	12		
		—	C	0	18		

[a]Post: posttest.
[b]W: wrong.
[c]C: correct.
[d]Pre: pretest.
[e]Not applicable.
[f]N/A: not applicable.

For the control group, McNemar tests were similarly conducted on the 25 items. Five items (#1, #4, #5, #18, and #19) showed statistically significant improvements ($P<.05$), though the gains were less pronounced than in the experimental group. Both groups improved significantly in aural recognition (#1, #4, and #5) and rhythmic analysis (#18 and #19). However, the experimental group showed greater gains in complex tasks like beat identification (#18) and score interpretation (#19), suggesting the GBL module excels at engaging pupils in challenging musical concepts. The improvements in the control group were more stable in simpler tasks (eg, #1 and #4), indicating retention rather than new learning, consistent with the limited scope of traditional approaches. The item-level analysis underscores that the GBL module improves piano music knowledge, particularly in rhythm and score analysis, where the experimental group outperformed the control group. Both groups mastered several things, but the larger and more substantial gains of the experimental group show that the GBL module can enhance music education outcomes.

Discussion

Summary of the Main Study

This study examined whether a GBL module for piano education could improve music knowledge among elementary school pupils in Anhui, China, compared with traditional piano instruction. The findings demonstrated that pupils in the gamified learning group significantly outperformed their peers in the control group after the 8-week intervention, achieving near-perfect posttest scores with reduced variability. This suggests that gamification not only facilitated mastery of theoretical and aural skills but also promoted more consistent learning outcomes across participants. The between-group Mann-Whitney *U* test showed a large effect size, while the within-group Wilcoxon analysis confirmed significant pre-post improvements in both groups, with stronger gains in the gamified condition. These results validate the hypothesis of the study that gamified learning would yield superior knowledge acquisition compared to conventional methods, addressing the gap identified in the Introduction. In summary, the evidence confirms that gamification is an effective and scalable pedagogical approach for strengthening music knowledge in early piano education.

A tailored GBL module improved elementary school pupils' piano music knowledge in Anhui, compared to the control group. Postintervention, GBL module students outperformed the control group by a significant margin. Gamified approaches improve music knowledge, supporting previous research on GBL for music education engagement [11,12]. The near-perfect scores of the experimental group and low variability indicate that the module boosted knowledge acquisition and ensured consistent mastery, a challenge in traditional piano pedagogy.

A comprehensive needs analysis indicated that Anhui elementary pupils struggle with motivation and foundational skills when learning piano. A dynamic, entertaining GBL approach was used to improve piano knowledge, skills, and attitudes. The module promoted musical competency and positive learning habits through gamification principles, including fun, clear objectives, immediate feedback, and adaptability [26]. Addressing these challenges created an engaging experience targeted to young learners' interests, establishing the groundwork for the research findings.

Starting with "Please Play" to build foundational skills and on to "Church Organ" to refine expressive techniques, the module material was progressive. Structured repetition improved cognitive and motor skills in this sequencing. The instructional design integrated traditional piano pedagogy with gamified activities like the "Right-Hand Shooting Game" and "Five-Line Speech Practice" to improve retention and motivation [27]. These entertaining activities improved rhythm, note recognition, and expression, directly benefiting the understanding of the experimental group.

The lesson design required only a piano and basic multimedia tools, ensuring accessibility across schools in Anhui. Teacher-led demonstrations emphasized real-time feedback and interpersonal connection, which young learners need, above digital media. Tailored feedback allowed pupils to correct errors and refine performance immediately [4], boosting motivation and addressing individual needs [27]. While technology provided aural and visual signals, research suggests that interleaving teacher demonstrations with student imitation is better than audio-only or blocked observation [28]. Thus, the module balanced traditional and technological approaches in a practical, learner-centered environment [28].

Comparisons to Existing Literature

Quasi-experimental assessments confirmed the efficacy of the module. The pupils improved their piano knowledge and skills from a poor baseline to mastery by the end of the study, outperforming peers in traditional settings. Item-level analysis demonstrated targeted gains in aural recognition (pitch differentiation) and rhythmic skills (beat identification), aligning with the gamified focus [29]. Game-based elements, including challenges, real-time feedback, and rewards, increased the median score of the experimental group much more than the control group. Similar to Molloy et al [21], gamified piano instruction enhanced note accuracy and technical performance compared to traditional techniques, possibly due to greater engagement and skill mastery. Fadhli et al [30] found in a meta-analysis that gamified instruction improves children's musical knowledge and technical proficiency. These findings support the idea that GBL approaches can significantly improve musical learning outcomes.

Both groups started with similar baselines, but the experimental group improved by posttest. Traditional teaching methods yielded marginal, inconsistent gains in the control group. GBL improves engagement and retention, according to Robert et al [31] and Qian and Jiang [32], who linked it to piano performance. The interactive design of the module addressed the focus of traditional pedagogy on mechanical execution over holistic understanding, resulting in limited development in the control group [33,34].

The success of the experimental group reflects the affective and cognitive benefits of GBL. Gamification increases motivation and self-efficacy, while narrative-driven GBL deepens emotional connections to information [35]. These mechanisms likely drove the experimental group to engage and master. Molero et al [36] observed that gamified systems simplify challenging concepts, which promotes integration of Orff and Dalcroze approaches [15,37]. The focus of the module on aural and rhythmic skills addressed music education gaps, where an outdated curriculum hinders creativity [32]. Beyond immediate outcomes, the performance of the experimental group in analytical tasks (eg, interval comparison) implies GBL fosters higher-order thinking and skill transferability, areas understudied [17,38].

Implementation fidelity was generally high, as teachers followed the structured lesson plans and completed weekly fidelity checklists. However, minor barriers were observed, particularly in maintaining pupil focus during transitions from game activities to reflective discussions. These challenges suggest that future iterations of the module should incorporate more structured debriefing routines to support smoother transitions and deepen conceptual understanding. This study builds on the Orff approach, Dalcroze Eurhythmics, and the SMDM to better understand how gamified learning improves music knowledge in elementary school pupils. The interactive, rhythm-focused exercises of the game-based module increased engagement and musical comprehension, following the Orff approach [13]. The kinesthetic features of the module follow Dalcroze Eurhythmics, which integrates movement and music to improve rhythmic and expressive skills [14], improving pupils' competency and sensitivity.

Practically, the validated game-based module offers Anhui educators a scalable, resource-efficient tool to modernize piano teaching in often underresourced settings. The experimental group outperformed the control group significantly in knowledge and skills underscores the efficacy of integrating game elements, such as points, levels, and immediate feedback, into instruction [4]. By fostering enjoyment and reducing anxiety, the module also sustains long-term engagement, addressing dropout risks and challenging traditional methods' focus on technical drills over emotional connection [39,40]. Curriculum designers should thus prioritize holistic, learner-centered approaches.

Limitations

Despite the promising outcomes of this study, several limitations warrant consideration to contextualize the findings and guide future research. First, the quasi-experimental design used a nonequivalent control group without true random assignment, which may have introduced selection bias, as evidenced by initial differences in pretest scores [24]. Future studies should use true random assignment to enhance internal validity and ensure baseline equivalence between groups. Additionally, the sample size of 60 participants, while sufficient for detecting significant effects, limits the generalizability of findings. The study focused on novice piano learners in a private training center in Anhui, which may not fully represent the diverse educational contexts across China or other regions. Expanding the sample size and including public school settings could strengthen the applicability of the GBL module.

The 8-week intervention period provided valuable insights into short-term knowledge gains but was insufficient to assess long-term retention or the sustained impact of GBL on musical development. Previous research highlights the need for longitudinal data to evaluate deep conceptual understanding and skill retention [10]. Future studies should extend the intervention duration and incorporate follow-up assessments to examine whether the observed gains persist over time. While gamification enhanced engagement, teachers occasionally reported difficulty redirecting pupils' attention from gameplay to reflective discussion. To mitigate this, future iterations may include structured debrief sessions after each activity.

Another limitation lies in the assessment tool, the Level 1 Basic Music Written Test, which exhibited ceiling effects in the experimental group through their posttest scores. These effects suggest that the test may not have been sufficiently challenging to fully capture knowledge gains, potentially underestimating the impact of the intervention. Moreover, the study primarily measures aural and theoretical knowledge. This focus may overlook other critical aspects of piano learning, such as technical proficiency, emotional expression, or creativity, which are integral to holistic music education [7]. Incorporating multifaceted assessment tools, such as performance evaluations or qualitative measures of student engagement, would provide a more comprehensive understanding of the impact of GBL. Future studies may supplement written tests with recorded piano performances assessed using standardized rubrics, teacher observations, and student portfolios, enabling a more holistic evaluation of musical skill development.

The study did not account for individual learner differences, such as cognitive abilities, learning styles, or previous musical exposure, which may influence the efficacy of GBL [41,42]. The GBL module's one-size-fits-all approach, while effective overall, may not equally benefit all pupils. For instance, item-level analysis revealed variability in performance on complex tasks like interval comparison, suggesting potential confusion for some learners. Future research should explore adaptive GBL designs that tailor content to individual needs. Additionally, the reliance on teacher-led instruction, while resource-efficient, may have introduced variability in implementation fidelity. Although teachers underwent training, differences in their delivery styles could have influenced outcomes.

Finally, this study focuses on a specific context of Anhui, while a strength for regional relevance, but limits its generalizability to other cultural or educational settings. Local teaching practices and resource constraints in Anhui shaped the design of the module, but these may differ elsewhere. Comparative studies across diverse regions could elucidate how contextual factors influence the effectiveness of the GBL module. Furthermore, its integration with traditional pedagogy like Orff and Dalcroze was effective, but its compatibility with other music education frameworks remains underexplored. Investigating hybrid models that combine GBL with varied pedagogical approaches could broaden its applicability.

To address these limitations, researchers should prioritize longitudinal studies with larger, more diverse samples to validate and extend the findings. Randomizing group assignments and incorporating varied assessment methods will enhance the robustness of future investigations. Educators in Anhui and similar regions are encouraged to adopt the validated GBL module, adapting it to local needs while ensuring teacher training emphasizes consistent implementation. Policymakers should consider investing in scalable GBL frameworks, potentially integrating digital tools to enhance accessibility and personalization in resource-constrained settings. By addressing these limitations and implementing these recommendations, the field can further harness the potential of GBL to transform piano education, fostering both musical proficiency and lifelong engagement among young learners.

Conclusion

The design and validation of the module underscore key implications for music education. Gamification enhanced engagement and knowledge acquisition, suggesting its potential for broader application. Multimodal instruction catered to diverse learning styles, while collaborative practices extended learning beyond the classroom. In conclusion, this GBL module provides a robust, validated approach to early music education for Anhui pupils. By integrating structured content, playful activities, and reflective teaching, it strengthens musical competencies and fosters a positive learning ecosystem, contributing to the evidence base for innovative pedagogy.

This study highlights that integrating gamification into piano pedagogy is both innovative and impactful for early music education. By embedding elements of play, challenge, feedback, and progression into traditional piano lessons, the GBL module transformed abstract musical concepts into engaging learning experiences that fostered consistent mastery among elementary school pupils. This approach is particularly valuable in resource-constrained environments, as the module was designed to be cost-effective, requiring minimal technology beyond a piano and simple teaching aids, while still producing large and meaningful learning gains. The innovation lies in bridging established pedagogical traditions, such as Orff and Dalcroze, with systematic gamification through the SMDM, thereby creating a hybrid framework that supports motivation, knowledge retention, and skill transfer. While traditional methods often emphasize technical drills at the expense of holistic understanding, this study shows that gamification sustains engagement and cultivates deeper conceptual knowledge. Practically, the validated GBL module can be adopted widely by music educators and curriculum designers to modernize piano teaching, reduce dropout risks, and increase accessibility of music learning across diverse school settings. Ultimately, the study demonstrates that gamified learning is not only a pedagogical enhancement but also a scalable and sustainable solution for advancing children’s cultural capital, creativity, and long-term musical engagement.

In conclusion, this study demonstrates that integrating gamification into elementary piano education represents both an innovative and effective pedagogical advancement. The study differs from existing research by systematically embedding GBL within established music education frameworks, namely Orff-Schulwerk and Dalcroze Eurhythmics, and operationalizing this integration through the SMDM. This approach moves beyond previous studies that primarily focus on engagement or short-term skill gains by providing robust empirical evidence of consistent knowledge mastery and reduced performance variability among young learners. The contribution of this study lies in offering a validated, theoretically grounded, and developmentally appropriate GBL module that advances understanding of how gamification can support foundational music knowledge acquisition. From a real-world perspective, the module provides educators and curriculum designers with a scalable, low-cost, and practical solution for modernizing piano instruction, particularly in underresourced educational contexts, while fostering sustained engagement, reducing learner attrition, and strengthening long-term musical development.

Data Availability

The data generated and analyzed during this study are not publicly available due to institutional and participant privacy considerations, but are available from the corresponding author upon reasonable request. Deidentified datasets, analysis code, and supplementary materials can be provided to qualified researchers for academic and noncommercial purposes.

Conflicts of Interest

None declared.

References

1. Bai B. Piano learning in the context of schooling during China's 'Piano craze' and beyond: motivations and pathways. Music Education Research 2021;23(4):512-526. [doi: 10.1080/14613808.2021.1929139]
2. Liu J. Compare the effect of different teaching methods in piano education. HBEM 2024;43:661-666. [doi: 10.54097/nt1nbh54]
3. Lin C, Liu EZ, Chen Y. Game-based remedial instruction in mastery learning for upper-primary school students. J Educ Tech Soc 2013:271-281 [FREE Full text]
4. Shi M. The exploration and application of children's music education in piano teaching. artpl 2023;4(7). [doi: 10.23977/artpl.2023.040707]
5. Sun Y. Comparative analysis of methods for teaching piano in various cultural traditions: from the classical European school to modern American approaches. Educ Manag Rev 2023;13 (12-1):191-197. [doi: 10.25726/j2371-4388-1702-x]
6. Garn AC, Jolly JL. High ability students' voice on learning motivation. J Advanced Academics 2013;25(1):7-24. [doi: 10.1177/1932202x13513262]
7. Reinsalu H, Timostsuk I, Ruokonen I. Literature review about the learning engagement in preschool and primary school music education. PMP 2022;21(2):35-69. [doi: 10.59893/pmp.21(2).004]
8. Yang L. Analysis of the current situation and improvement strategies for piano beginner teaching methods. Int J Educ Humanities 2025;20(3):9-12. [doi: 10.54097/rbn4f060]
9. Alotaibi MS. Game-based learning in early childhood education: a systematic review and meta-analysis. Front Psychol 2024;15:1307881 [FREE Full text] [doi: 10.3389/fpsyg.2024.1307881] [Medline: 38629045]
10. Weatherly KICH, Wright B, Lee EY. Digital game–based learning in music education: a systematic review between 2011 and 2023. Res Studies Music Educ 2024. [doi: 10.1177/1321103x241270819]
11. Liu Y. The application of music games in elementary school classes. Int J Educ Humanities 2023;9(2):205-208. [doi: 10.54097/ijeh.v9i2.9934]
12. Phunsa S, Pawala T. Learning through gameplay: creating memorable learning experiences with games. J Educ Learn Rev 2024;1(1):39-46. [doi: 10.60027/jelr.2024.678]
13. Orff C. The Schulwerk: its origin and aims. Music Educ J 1963;49(5):69-74. [doi: 10.2307/3389951]
14. Jaques-Dalcroze, Rubinstein H. Rhythm, Music and Education. London: Chatto & Windus; 1921.
15. Beegle A. Playing with Orff and classroom instruments. In: Abril CR, Gault BM, editors. General Music: Dimensions of Practice. New York, NY: Oxford University Press; 2022.
16. Daly DK. Creativity, autonomy and Dalcroze eurhythmics: an arts practice exploration. Int J Music Educ 2021;40(1):105-117. [doi: 10.1177/02557614211028600]
17. Liu L. Research on piano education in contemporary china. J Contemp Educ Res 2021;5(2). [doi: 10.26689/jcer.v5i2.1836]
18. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004;94(3):361-366. [doi: 10.2105/ajph.94.3.361] [Medline: 14998794]
19. Sidek MN, Jamaludin A. Pembinaan modul: bagaimana membina modul latihan dan modul akademik. Serdang: Penerbit Universiti; 2005.
20. Madihie A, Noah SM. An application of the Sidek module development in rebt counseling intervention module design for orphans. Procedia - Social and Behavioral Sciences 2013;84:1481-1491. [doi: 10.1016/j.sbspro.2013.06.777]
21. Molloy W, Huang E, Wünsche B. Mixed reality piano tutor: a gamified piano practice environment. 2019 Presented at: Proceedings of the International Conference on Electronics, Information, and Communication (ICEIC); May 6, 2019; Auckland, New Zealand p. 22-25. [doi: 10.23919/elinfocom.2019.8706474]
22. Lazzaro N. The 4 keys 2 fun. XEODesign. URL: http://www.xeodesign.com/research/ [accessed 2025-06-28]
23. Mohammed J, Schmidt K, Williams J. Designing a new course using backward design. 2022 Presented at: Proceedings of the ASEE Annual Conference & Exposition; June 26, 2022; Minneapolis, USA p. 26-29. [doi: 10.18260/1-2--41806]
24. Hinton PR. Mann-Whitney U test. In: Salkind NJ, editor. The SAGE Encyclopedia of Social Science Research Methods. Thousand Oaks, CA: SAGE Publications; 2010:748.
25. Hupe M. StatPearls. J Electron Res Med Libraries 2024;21(3):147-152. [doi: 10.1080/15424065.2024.2389325]
26. Purinton EF, Burke MM. Engaging online students: using a multisensory exercise for deeper, active learning. Marketing Educ Rev 2019;30(1):29-42. [doi: 10.1080/10528008.2019.1677478]
27. Taran I. Musical activities and games in the process of learning to play the piano. Musical art in the educological discourse 2024(8):36-40. [doi: 10.28925/2518-766x.2023.86]
28. Simones L, Rodger M, Schroeder F. Seeing how it sounds: observation, imitation, and improved learning in piano playing. Cognition and Instruction 2017;35(2):125-140. [doi: 10.1080/07370008.2017.1282483]

29. Yang Z. Modern piano teaching technologies: accessibility, effectiveness, the need for pedagogues. Elementary Educ Online 2020;19(3):1812-1819 [FREE Full text]
30. Fadhli M, Brick B, Setyosari P, Ulfa S, Kuswandi D. A meta-analysis of selected studies on the effectiveness of gamification method for children. Int J Instruction 2020;13(1):845-854. [doi: 10.29333/iji.2020.13154a]
31. Robert D, Bt. Jamri N, Sandra Hazel Ling, Bt. Amin AA, Bt. Yazid FA. Gamified learning intervention to promote music literacy and creativity in elementary music education. JCSHD 2023;9(1):18-41. [doi: 10.33736/jcshd.5481.2023]
32. Qian C, Jiang M. Exploring the effects of digital game-based learning on music education. Studies in Soc Sci Humanities. URL: https://www.paradigmpress.org/SSSH/article/view/1118 [accessed 2025-12-23]
33. Smith S. Expressive piano technic: basics of healthy movement, technic and interpretation. Rast Müzikoloji Dergisi. 2021 2021;9(2):2757-2768. [doi: 10.12975/rastmd.2021923]
34. Martinović Bogojević J, Rotar Pance B. Musical creativity in the teaching practice in Montenegrin and Slovenian primary schools. Brit J Music Ed 2022;39(2):169-182. [doi: 10.1017/s0265051722000018]
35. Yeboah R, Amponsah KD, Commey Mintah P, Sedofia J, Kwarteng Donkor (Opare) PB. Game-based learning in Ghanaian primary schools: listening to the views of teachers. Education 3-13 2023;53(1):99-113. [doi: 10.1080/03004279.2023.2171269]
36. Molero D, Schez-Sobrino S, Vallejo D, Glez-Morcillo C, Albusac J. A novel approach to learning music and piano based on mixed reality and gamification. Multimed Tools Appl 2020;80(1):165-186. [doi: 10.1007/s11042-020-09678-9]
37. Guan Y, Meeson P. The application of Dalcroze eurhythmics in primary school music education. Educ Rev 2024;8(2):340-344. [doi: 10.26855/er.2024.02.027]
38. Safian A, Nasrifan MN. Higher-order thinking skills in music education: a comprehensive analysis. Grenek: j seni musik 2013;13(2):182-191. [doi: 10.24114/grenek.v13i2.62780]
39. Du S. What teaching methods can be used by piano teachers to improve learning outcomes of child and adult beginners? 2022 Presented at: Proceedings of the International Conference on Education, Language and Art; Nov 26-28, 2022; Sanya, China p. 26-28. [doi: 10.2991/assehr.k.220131.020]
40. Ou C. Analysis and comparison of methods of teaching piano art in classical and modern educational systems: from traditions to innovations [Article in Russian]. Educ Manag Rev 2024;14(1-2):256-264. [doi: 10.25726/y5657-0174-6659-o]
41. Ningtyas DP, Setyosari P, Kuswandi D, Ulfa S. Enhancing early childhood problem-solving abilities through game-based learning and computational thinking: the impact of cognitive styles. Golden Age SCI J ECD 2024;9(3):409-419. [doi: 10.14421/jga.2024.93-04]
42. Avcı A. Piano education through educational games: piano in the box game. Base for Electronic Educ Sci 2020;1(1):27-39. [doi: 10.29329/bedu.2020.253.3]

Abbreviations

GBL: game-based learning
MCAR: missing completely at random
SMDM: Sidek Module Development Model
TREND: Transparent Reporting of Evaluations with Nonrandomized Designs

Edited by S Brini; submitted 16.Jul.2025; peer-reviewed by H Haruna, S Papadakis, Y Asada; comments to author 05.Oct.2025; revised version received 15.Dec.2025; accepted 21.Dec.2025; published 19.Jan.2026.

Please cite as:
Wang Y, Tan WH, Ye Q, Gu T
Effects of Game-Based Learning on Piano Music Knowledge Among Elementary School Pupils: Pretest-Posttest Quasi-Experimental Study
JMIR Serious Games 2026;14:e80766
URL: https://games.jmir.org/2026/1/e80766
doi:10.2196/80766
PMID:

Original Paper

A Haptic-Driven Serious Game for Cognitive Stimulation and Visual Impairment Mitigation in Older Adults Based on the Design-Play-Experience Framework: Cross-Sectional Mixed Methods Pilot Study

Xin Huang[1], MA; Nazlena Mohamad Ali[1], PhD; Shafrida Sahrani[1], PhD; Yue Zhang[1], MA

Institute of Visual Informatics (IVI), Universiti Kebangsaan Malaysia, Selangor, Malaysia

Corresponding Author:
Nazlena Mohamad Ali, PhD
Institute of Visual Informatics (IVI)
Universiti Kebangsaan Malaysia
Aras 1, Akademia Siber Teknopolis, Universiti Kebangsaan Malaysia
Selangor, 43600
Malaysia
Phone: 60 389272402 ext 2425
Fax: 60 389272402
Email: nazlena.ali@ukm.edu.my

Abstract

Background: In the context of global aging, cognitive decline among older adults has become a prevalent issue, significantly impacting their daily lives. Serious games have demonstrated potential in enhancing cognitive abilities in this population. However, most existing serious games designed for older adults rely heavily on visual interfaces, which are often potentially detrimental for those with pre-existing visual impairments.

Objective: This study had two primary objectives: (1) to design a theoretical prototype for a haptic-driven serious game for older adults based on the Design-Play-Experience (DPE) framework, aiming to enhance cognitive abilities, including attention, logical reasoning, and decision-making while simultaneously mitigating challenges associated with visual impairment, and (2) to conduct a pilot study evaluating the prototype's usability, accessibility, and user experience within the target population.

Methods: We used a cross-sectional, mixed methods pilot study with a single-group observational design, comprising a theoretical design and a pilot user study. First, the DPE framework was systematically applied to develop a game prototype by integrating haptic feedback technology (using built-in smartphone vibration motors) across its 3 core dimensions: design (haptic symbol system, accessible interface), play (dynamic difficulty adjustment), and experience (emotional engagement). Subsequently, a pilot study was conducted with 10 older adults recruited via convenience sampling (mean age 62.9, SD 3.35 years; 5 male, 5 female; all with self-reported mild visual impairments, such as presbyopia). Following interaction with the prototype, data were collected remotely using the System Usability Scale (SUS) and semistructured interviews administered via videoconferencing. Quantitative data from the SUS were analyzed using descriptive statistics, while qualitative data from the interviews were processed using thematic analysis.

Results: Pilot user studies showed that the game prototype had good usability, with an average SUS score of 89.5 (SD 2.72; 95% CI 87.6-91.4), which is considered "excellent." Thematic analysis of the interviews revealed three significant themes. The first theme was intuitive haptic feedback, which reflected that participants were able to quickly grasp and value the vibrational cues used to identify cards. The second theme was based on reduced eye strain, in which the combination of large fonts, high-contrast interfaces, and haptic feedback was praised for its effectiveness in relieving eye strain. The third theme was simplicity, where the simplified card game mechanics were considered both fun and challenging.

Conclusions: This study developed and validated a haptic, serious game for older adults. Its innovation lies in the systematic application of the DPE framework to achieve "haptic substitution for vision," which differs from previous research that focused on general immersive experiences. The main contribution of this study is providing a reusable design blueprint for creating easy-to-use cognitive training tools. These findings have practical implications in the real world, providing a feasible approach for deploying low visual load interventions in communities and care facilities.

(JMIR Serious Games 2026;14:e86290) doi:10.2196/86290

KEYWORDS

serious games; haptic feedback; older adults; cognitive abilities; visual impairment; DPE framework; Design-Play-Experience; pilot study

Introduction

By 2050, two-thirds of the global population aged 60 years and older will live in low- and middle-income countries [1]. The population structure is undergoing a significant aging transition, and the most common cognitive diseases in older adults include mild cognitive impairment [2], dementia [3], and Parkinson disease [4]. Mild cognitive impairment is a cognitive disorder between dementia and health [5]. At this time, the patient's memory and cognitive function have already had problems, and one or more cognitive functions have declined, but the patient's daily life will not be significantly affected [6,7]. Dementia affects 36 million people worldwide [8]. Dementia is closely associated with age (aging), with nearly half of the population aged 85 years and older developing dementia [9]. The most common type of dementia is Alzheimer disease. The typical initial symptom is memory loss. Parkinson disease usually occurs in people aged 60 years and older. It is a chronic neurodegenerative disease that affects the central nervous system, with the most obvious early symptoms being tremors, limb stiffness, decreased movement, and gait abnormalities [10]. Parkinson disease may also cause cognitive and behavioral problems.

Serious games have had some success in helping older adults recover cognitive abilities [11]. Serious games are a set of solutions designed to make a range of rehabilitation courses more engaging and less boring [12]. MINWii (Samuel Benveniste, Pierre Jouvelot, and Renaud Pequignot) is a new serious video game tailored for people with Alzheimer disease and dementia, in which older gamers use Wiimotes to improvise or play predefined songs on a virtual keyboard displayed on the screen [13]. However, vision impairments in older adults may affect their interaction with serious games [14]. Older adults often experience sensory changes that affect their interaction with digital displays and games. These changes include decreased vision, impaired dark adaptation, reduced contrast sensitivity, limited visual accommodation range, reduced color sensitivity, and increased sensitivity to glare [14]. These impairments can make it difficult for older adults to perceive small elements on a screen, read text, or navigate complex interfaces [15]. The older adults said that the mobile phone screen was too small when interacting, which shows that the limitations of digital tools themselves are also one of the factors that affect the older adults' interaction with serious games [16]. These studies show that, in addition to the decline of the older adult's own sensory organs over time, serious games designed for older adults do not fully consider the older adult's user experience.

The addition of haptic feedback technology further expands the impact of serious games [17] and can reduce the visual damage caused by long-term serious game playing by older adults. The haptic experience provides a new way of perception for older adults, thereby enhancing the user's cognitive ability. Haptic technology is a sensory feedback technology that simulates haptic perception through force, vibration, or motion. It captures the user's movements and interaction information through sensors and generates corresponding haptic feedback through actuators, allowing users to feel real physical interactions in a virtual environment [17]. Haptic feedback technology provides an immersive experience for older adults by combining their sense of touch with what they see, hear, or interact with [18]. At the same time, serious games combined with haptic feedback technology have shown the potential to significantly improve cognitive abilities [19]. For example, in mobile serious games, the interaction method developed by Deng et al [20] that combines eye tracking with haptic feedback also significantly improves the user's operation accuracy and task completion rate. By providing haptic feedback on mobile devices, the user's error rate in interactive tasks is significantly reduced, indicating that serious games combined with haptic feedback can effectively enhance the user's understanding and control of cognitive abilities [20]. Silva et al [21] reported a new haptic device under development, mainly for older adults, to stimulate and quantify the response of the older adults' nervous system through serious games.

The existing combination of serious games and haptic feedback technology mainly focuses on enhancing interactive immersion. Although it has improved the cognitive ability of some older people, there are few system frameworks specifically for older adults with visual impairment. There is a lack of deep integration of cognitive training and haptic feedback. Its design is mostly oriented to general users and lacks targeted adaptation for the visual and haptic coordination needs of older adults. There are 2 major limitations in the current research on the use of serious games by older adults, one being the problem of visual dependence. Most serious games rely on complex visual interfaces, which aggravate the interaction barriers caused by vision decline in older adults. The second is the one-sidedness of haptic feedback design. Although haptic technology can enhance immersion through force feedback, its application focuses on general scenarios and lacks a system framework for the cognitive decline and sensory coordination needs of older adults (such as "haptic substitution for vision").

To address these shortcomings, this study uses the Design-Play-Experience (DPE) framework [22] as its theoretical basis to develop a prototype of a haptic-driven serious card game. The DPE framework provides a structured approach to integrate technology, game mechanics, and user experience, making it well-suited for systematically incorporating haptic feedback to achieve specific cognitive and accessibility goals. The final prototype embodies this integration (1) in the design dimension, through a haptic symbol system and a high contrast interface; (2) in the play dimension, through haptic cues to dynamically adjust the game difficulty; and (3) in the experience dimension, through haptic rewards to enhance emotional engagement. The objective of this study is to develop a haptic feedback serious card game theoretical prototype based on the DPE framework to reduce the visual burden of older adults and improve their cognitive ability. Second, to conduct a pilot study evaluating the prototype's usability, accessibility, and user experience within the target population. The contribution of this study is to combine haptic feedback with the card game mechanism and provide a reusable design blueprint for the design of older adult–friendly serious games through the system mapping of the DPE framework.

Methods

Quantitative Analysis

Inclusion and Exclusion

The inclusion criteria for participants were (1) age ≥60 years; (2) self-reported mild visual impairment, such as presbyopia; (3) basic smartphone usage skills; and (4) willingness to sign an informed consent form. The exclusion criteria included (1) being diagnosed with severe cognitive impairment, such as clinically diagnosed dementia, and (2) having severe hearing loss or motor impairment that significantly affects haptic perception or device operation.

Participant Characteristics

Participants had a mean age of 62.9 (SD 3.35) years, with an age range of 60-71 years, and comprised 5 males and 5 females. All participants reported having mild visual impairment, such as presbyopia, and possessed basic smartphone operation skills.

Sampling Procedures

Participants were recruited from Dangtu Old Age University, using convenience sampling [23]. Recruitment and data collection were conducted remotely from June 1 to July 1, 2025. The research team also collaborated with the institution's administration to disseminate recruitment information through employee referrals. Interested potential seniors contacted the research team by phone, and those who passed the initial screening were formally enrolled.

Sample Size, Power, and Precision

As a pilot study, the sample size (N=10) was determined primarily based on feasibility considerations rather than statistical power calculations. This sample size is typical for pilot studies and aims to provide preliminary data and process validation for subsequent large-scale efficacy trials.

Measures and Covariates

The primary outcome measure was usability of the game prototype, measured using the System Usability Scale (SUS). The scale's total score ranges from 0 to 100. The secondary outcome measure includes accessibility and user experience. In covariates and confounding factors, participants' baseline cognitive level, severity of visual impairment, technical experience, and age were recorded as potential confounding factors and qualitatively documented during interviews via background questions.

Data Collection

Quantitative data were collected through online survey platforms, such as Wenjuanxing [24]. Participants completed a SUS [25] online after interacting with the game prototype.

Quality of Measurements

To ensure data quality, all data collection procedures were standardized. Specifically, all participants used the same game prototype and completed the same SUS. Furthermore, a consistent interview process was followed, with semistructured interviews [26] conducted by the same researcher after participants had finished experiencing the game, using interview guidelines to ensure consistency across all interviews. The SUS [25], the primary quantitative tool in this study, has had its reliability and validity extensively confirmed in previous research [27,28].

Instrumentation

The main tools used in this study were the SUS and a haptic game prototype developed based on the DPE framework. This prototype runs on an Android smartphone and uses the device's built-in linear resonant motor to provide structured haptic feedback.

Masking

Given that this study used a single-group observational design aimed at directly obtaining users' subjective feedback on the prototype, no masking was applied to participants or researchers.

Psychometrics

This study used the widely validated SUS, whose reliability and validity are supported by literature [27]. SUS has demonstrated excellent psychometric properties, including good reliability (eg, high internal consistency, Cronbach alpha coefficient typically above 0.85, and test-retest reliability of approximately 0.80) and strong validity (convergent validity, discriminant validity, and construct validity) [25,27,28]. This study directly reports the SUS scores observed in the current sample.

Conditions and Design

This study uses a nonexperimental observational design, specifically a single-group, cross-sectional study. Reporting of the study was performed in accordance with the JARS (Journal Article Reporting Standards) guidelines [29].

Data Diagnostics

The study conducted diagnostic checks on the data. All 10 participants completed the entire study process, and there were no missing data for either the primary or secondary outcome measures. Therefore, no imputation was performed. Given the small sample size, the distribution of SUS scores was examined. No extreme outliers requiring intervention were found, and no data transformation was performed.

Analytic Strategy

Quantitative data analysis used descriptive statistics. The total score of the SUS was described using the mean and SD, and 95% CIs based on a *t* test distribution were calculated to provide more conservative interval estimates for small samples. All analyses aimed to assess the initial usability of the game prototype.

Qualitative Analysis

Research Design Overview

The qualitative portion of this study used a thematic analysis [30] approach to explore older users' experiences, perceived accessibility, and emotional engagement with haptic-driven game prototypes. This design was well-suited for gathering detailed participant-centered perspectives from semistructured interviews.

Study Participants or Data Sources

The qualitative data came from the same 10 participants (mean age of 62.9, SD 3.35 years; 5 male and 5 female) as the quantitative data. The data were in the form of audio recordings of semistructured interviews with each participant and their verbatim transcripts.

Participant Recruitment

The recruitment process for participants was the same as that for the quantitative part. Convenience sampling was used to recruit from Dangtu Old Age University. The research team collaborated with the institution to make initial contact, and eligible volunteers were enrolled after confirmation by phone. All participants in the quantitative assessment completed subsequent qualitative interviews to ensure the continuity of data sources.

Data Collection

This study collected qualitative data through semistructured interviews conducted using videoconferencing software, such as Tencent Meeting [31]. We developed an interview guideline to explore participants' experiences in three areas: (1) the intuitiveness of haptic feedback, (2) the impact of haptic feedback on alleviating visual fatigue, and (3) the overall appeal and engagement of the game. This approach ensured comprehensive data collection. All interviews were conducted immediately after participants completed the game experience and SUSs to ensure the freshness of the experience. Data collection continued until thematic saturation was reached. All interviews were recorded with the participants' consent and transcribed by researchers for analysis. All interviews were recorded with the participants' consent and transcribed by researchers for analysis.

Ethical Considerations

The study was approved by the ethics committee before it was carried out, and it strictly adhered to the ethical guidelines for research involving human participants. The study protocol, including the data collection methods, was considered to comply with the ethical standards for research involving human participants (approval reference UKM.IVI.600/8/1-P136397). Written informed consent was obtained from all participants before participating in the study. The consent form clearly outlined the purpose, procedures, risks, benefits, and voluntary nature of participating in the study. Participants were informed of their right to withdraw at any time without penalty. For secondary data analysis, the original consent form explicitly allowed the use of deidentified data for research purposes without the need for additional consent. All participant data were anonymized and deidentified during collection and analysis. Personally identifiable information was removed from survey responses, and data were securely stored on a password-protected server accessible only to the research team. No identifiable information will be shared with external parties or institutions. Participants received no monetary or nonmonetary compensation for their participation in this study. Their contribution was voluntary and motivated by the potential societal benefits of the research. No identification of individual participants in any images of the paper or supplementary material is possible. All data reported were aggregated and anonymized to ensure participant confidentiality. This study conformed to the tenets of the Declaration of Helsinki and adhered to the ethical standards outlined by JMIR for research involving human participants.

Results

Participant Flow and Demographics

All 10 recruited older participants (5 male and 5 female; mean age 62.9, SD 3.35; 60-71 years) completed the entire research process, including interactive games, the SUS questionnaire, and interviews. No data were missing for either the primary or secondary outcome measures.

Quantitative Analysis

Quantitative analysis of the pilot study showed that the haptic game prototype received a very high usability rating. The mean score on the SUS was 89.5 (SD 2.72; 95% CI 87.6-91.4). According to the scale's evaluation criteria, this score falls into the "excellent" category, clearly indicating that the prototype possesses both high usability and high acceptability among the target older adults (Table 1).

Table 1. System Usability Scale (SUS) scores from a cross-sectional pilot study evaluating a haptic-driven serious game prototype for cognitive stimulation and visual impairment mitigation in older adults.

Metric	Value
SUS score, mean (SD; 95% CI)	89.5 (2.72; 87.6-91.4)
Adjective rating	Excellent
Score range	85.0-92.5

Qualitative Analysis

After conducting a thematic analysis of interviews, researchers identified three prominent themes regarding user experience and usability:

1. Intuitive haptic feedback: Participants quickly grasped and highly valued the vibrational cues used to identify cards. This haptic feedback was considered a natural and effective alternative to constant visual confirmation. One participant stated, "I quickly learned what the different vibrations represented; I didn't have to constantly stare at the screen to remember what cards I was holding."
2. Effective reduction of eye strain: The combination of haptic feedback and a high-contrast, large-font visual interface significantly reduced eye strain and made gameplay more comfortable. One participant commented, "The large font and vibrational cues combined made my eyes much more comfortable."
3. Simplicity: The simplified game mechanics (a simplified 7-card Doudizhu) were considered both fun and challenging, lowering the learning curve while maintaining player engagement through a dynamic difficulty adjustment system. One participant said, "It's easy to pick up, but still requires thought. It's fun without being frustrating."

A Theoretical Prototype of Serious Card Games Based on the DPE Framework

Table 2 is a mapping table of the DPE framework in game design. It comprehensively elaborates on the DPE framework for serious games aimed at older adults from 3 vertical dimensions (design, play, and experience) and 4 horizontal dimensions (learning, storytelling, gameplay, and user experience). The paper designs a theoretical prototype of a serious game with haptic feedback, aiming to improve the cognitive abilities of older adults and alleviate their visual impairments. In the design dimension, we created a haptic symbol system to reduce reliance on vision. In the gameplay dimension, we designed haptic cues to facilitate dynamic adjustment of game difficulty, and in the experience dimension, we added haptic rewards to enhance the player's emotional engagement.

Table 2. Prototype structure mapping for a haptic-driven serious card game based on the Design-Play-Experience (DPE) framework.

Core Pillar	Design	Play	Experience
Learning	• Content: Target cognitive abilities, such as attention, thinking, decision-making, and logical reasoning. • Teaching methods: digital cognition, simple mathematical operations, and decision-making.	• Teaching: Roles: Players play community members and play card games with virtual opponents. • Setting: Community environment, build the community by winning games. • Narrative: Earn coins by winning games, and build the community until the community is complete.	• Learning: Through gaming activities, older adults can practice and improve their cognitive abilities. • Self-efficacy: By successfully completing tasks through gaming, older adults can enhance their sense of self-efficacy and improve their self-confidence.
Storytelling	• Roles: Players play community members and play card games with virtual opponents. • Setting: Community environment where players build the community by winning games. • Narrative: Earn coins by winning games, and build the community until the community is complete.	• Storytelling: Players win coins and build communities by winning games. • Interactive narrative: The storyline can have different development paths based on the player's choices and performance.	• Story: Provide a rich story experience through the game storyline to increase the appeal of the game.
Gameplay	• Mechanics: A simplified version of Doudizhu, with only 7 cards, smaller than playing cards. Haptic feedback prompts, identifying the size of the card based on the number of vibrations (card 1, slight vibration 1 time, and so on).	• Dynamics: A smooth gameplay flow ensures a cohesive and engaging experience. • Difficulty Adjustment: Haptic feedback automatically adjusts the game difficulty based on player performance, maintaining the game's challenge and fun.	• Emotion: Design game elements that evoke positive emotional experiences, such as a sense of accomplishment and satisfaction. Provide unique haptic feedback as a reward when players achieve something or win, enhancing emotional engagement and a sense of accomplishment.
User experience	• User interface: Large fonts, high contrast, and a simple operation interface are suitable for older adults. • Navigation: Clear navigation and instructions help older adults easily find the required functions.	• Interactivity: Encourage players to interact with virtual opponents through game tasks. • Feedback interaction: Players' actions will receive immediate feedback, enhancing their sense of participation.	• Engagement: Design game elements that can attract players to continue to participate, such as achievement systems, reward mechanisms, etc. Enhance players' sense of belonging and community through the game community.

Game Prototype Interface Display

This study follows a prototype based on the DPE framework and systematically develops the visual and interactive design of the "Old Friends" game. We follow the WCAG 2.1 [32] accessibility standard and are explicitly committed to minimizing the visual and cognitive burden on older users throughout the design process.

The interface design intentionally used large, clear, and legible fonts (18 pt) and high contrast (4.5:1), consistently applied to all core visual components, including the game logo, card elements, login paper, main game, win or lose, and homeland interface. These choices directly supported the user experience goal emphasized in the design framework, focusing on clarity and usability. A key outcome of this process was the "Old Friends" logo (Figure 1). Its design exemplifies how accessibility principles translate into tangible visual identity. We chose the Bauhaus 93 font for its high legibility and clear glyphs, and simultaneously, we used a simple black-on-white color scheme to ensure maximum readability in accordance with the "minimum visual load principle." However, this functional clarity is complemented by carefully designed emotional elements. The name "Old Friends" and its warm, friendly presentation aim to resonate with the project's core objectives, emotional support, and social interaction, as detailed in our "Storytelling" and "Experience" dimensions. We believe this semantic and visual consistency enhances the emotional connection between users and the game, thereby increasing user engagement and strengthening community awareness. Therefore, the logo is not merely an aesthetic element but a reflection of a design philosophy that perfectly blends technological usability with an empathetic, goal-oriented narrative design.

Figure 1. Logo of the "Old Friends" serious card game, compliant with WCAG (Web Content Accessibility Guidelines) 2.1 standards and embodying high-contrast visual identity principles.

Figure 2 shows the final card design, developed through iterative iterations while consistently prioritizing senior friendliness. A core design decision was to use the clear and legible Berlin Sans Demi font to represent the numbers 1-7 at an oversized 120-point font size to ensure quick recognition. This design, combined with a high-contrast black-on-white scheme, forms the functional foundation of the card. Subsequently, the designers consciously layered various aesthetic elements on top of this easily understood foundation. Light blue and medium yellow circular decorations were introduced, adding visual interest and warmth without compromising clarity. The "Old Friends" logo, with its coordinated typography, visually and thematically unifies the entire component. This fusion of clear functionality and warm aesthetics was a thoughtful choice, designed to directly support the research goals of "emotional support" and "social interaction" by making the interface more appealing and personal for older users.

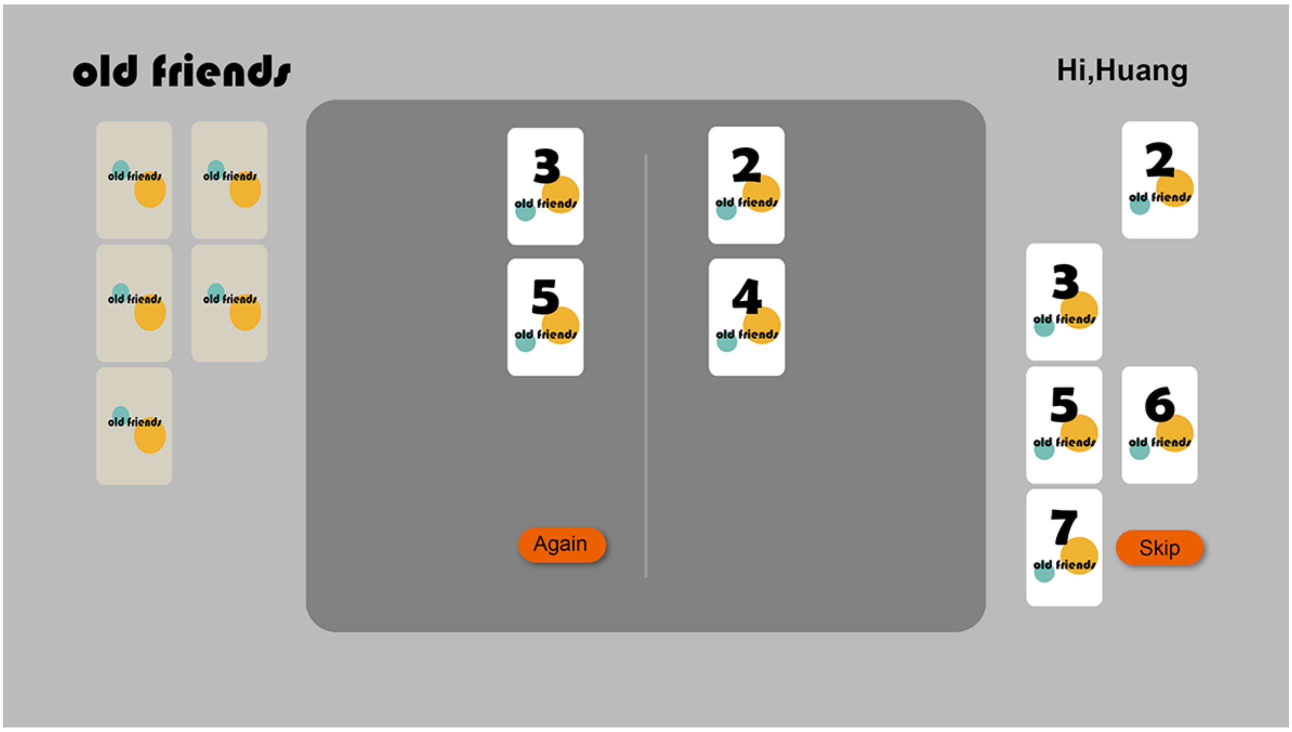
Figure 2. Game card interface of the “Old Friends” serious card game.

The login screen (Figure 3) is designed to immediately create a user-friendly atmosphere. To this end, we used a soft, light background, accented with soft yellow, orange, and blue circular graphic elements. This color scheme was deliberately chosen to create a warm and vibrant atmosphere without being overly stimulating to the senses, directly adhering to the “minimum visual load principle” and minimizing strong contrasts in unnecessary areas. Functional clarity is paramount. The central title, “Old Friends,” uses a high-contrast 60-point Bauhaus 93 font to ensure users can immediately recognize the brand. Similarly, the key operation buttons at the bottom of the screen (Start, Register, Homeland, Archives) use a standard, clear, and legible 18-point Arial font. They are arranged with ample spacing on a white background, a carefully designed interaction method intended to prevent accidental clicks and guide users easily into the game. Beyond functionality, the illustration of older people playing card games in the background also carries significant narrative meaning. It was included to visually convey the core themes of “emotional support” and “social interaction” from the outset, preparing users for a positive social experience and fostering a direct emotional connection with the game’s objectives.

Figure 3. Login page interface of the “Old Friends” serious card game.

The design of the game’s main interface (Figure 4) continues the aesthetic and functional principles established on the login page, aiming to create a low-stimulation environment for extended gameplay. We chose a neutral gray background, forming a calm and professional base color scheme that effectively reduces visual fatigue, adhering to the “minimum visual load principle.” Against this background, key interactive components, such as white cards and orange function buttons, are presented with high contrast to ensure immediate visual recognition and readability. The interface layout is carefully planned to guide the user’s interaction flow, with the left panel displaying the opponent’s hand, while the right area is dedicated to displaying the user’s hand, creating a clear and intuitive game cycle. The “Again” and “Skip” buttons at the bottom provide core operational controls, with their orange color ensuring high visibility. We clearly defined their functions, “Again” for haptic recognition of both players’ cards again, and “Skip” for skipping the current round, thus supporting the user’s cognitive rhythm and giving them clear operational autonomy. Emotional design remains indispensable at this stage. The consistently displayed “Old Friends” branding and personalized welcome messages, such as “Hello, Mr. Huang,” are not mere decorations. They are thoughtfully designed features to reinforce a sense of familiarity and social presence. This consistent emotional framework is crucial for aligning the core gameplay with the overall research objectives of “emotional support” and “social interaction.”

Figure 4. Main gameplay interface of the “Old Friends” serious card game.



The core design of the victory and defeat interfaces (Figures 5 and 6) is to provide emotionally intelligent feedback for older users. Both interfaces strictly adhere to the design language of the main interface, using a low-visual-load gray background and high-contrast text to ensure absolute clarity of information. The key design difference lies in the refined tuning of emotional semantics. The victory interface uses strongly affirmative language, such as “YOU ARE THE BEST!” to maximize the user’s sense of accomplishment and self-efficacy, while the defeat interface uses encouraging language, such as “The next one will be better!” to buffer frustration and guide the outcome toward a positive and constructive direction. This differentiated copywriting strategy is a conscious design choice to achieve the goal of “emotional support.” At the interaction level, the “Again” button is placed in the exact same position on both interfaces, following the principle of consistency. This allows users to quickly start a new game without having to search for it again, regardless of the outcome, reducing cognitive load and operational hesitation. The brand logo and personalized greeting (“Hi, Huang”) at the top serve as a consistent emotional thread, continuously reinforcing the game’s social sense of belonging. Overall, these 2 interfaces are not only the end point of the game process but also key nodes that connect the beginning and the end, maintaining the user’s emotional investment and continued participation.

Figure 5. Victory outcome interface of the “Old Friends” serious card game.

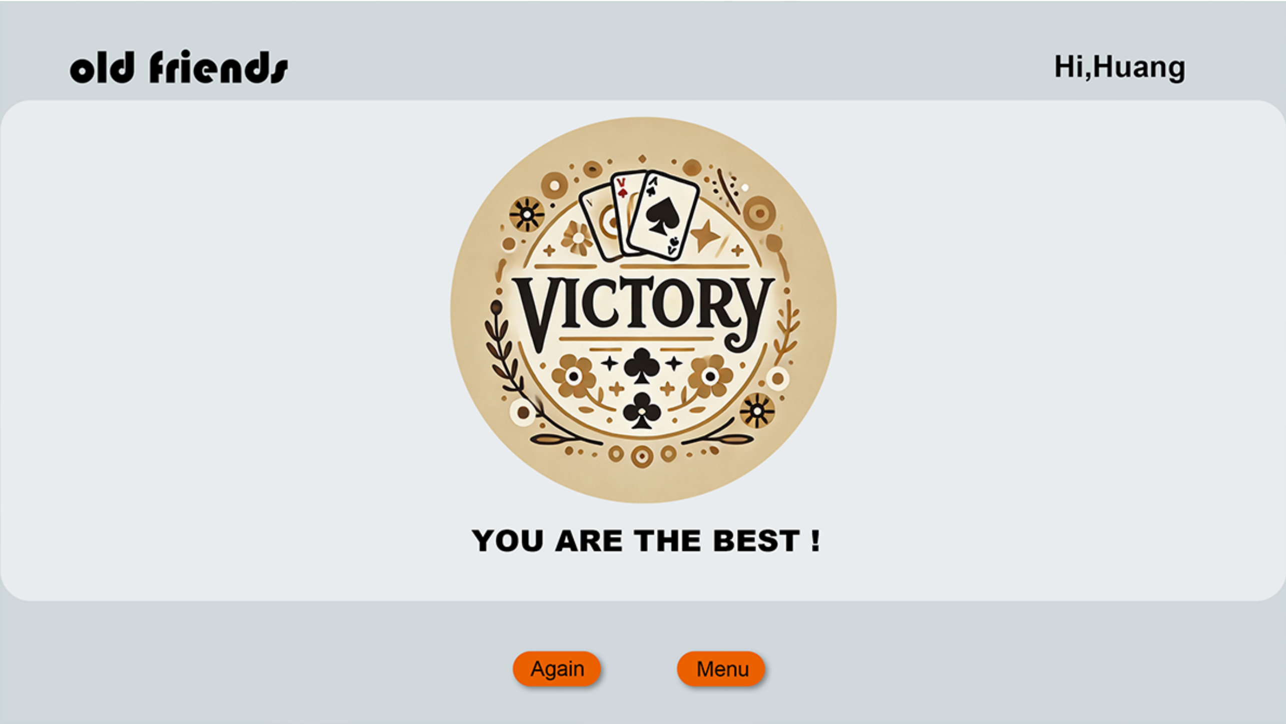
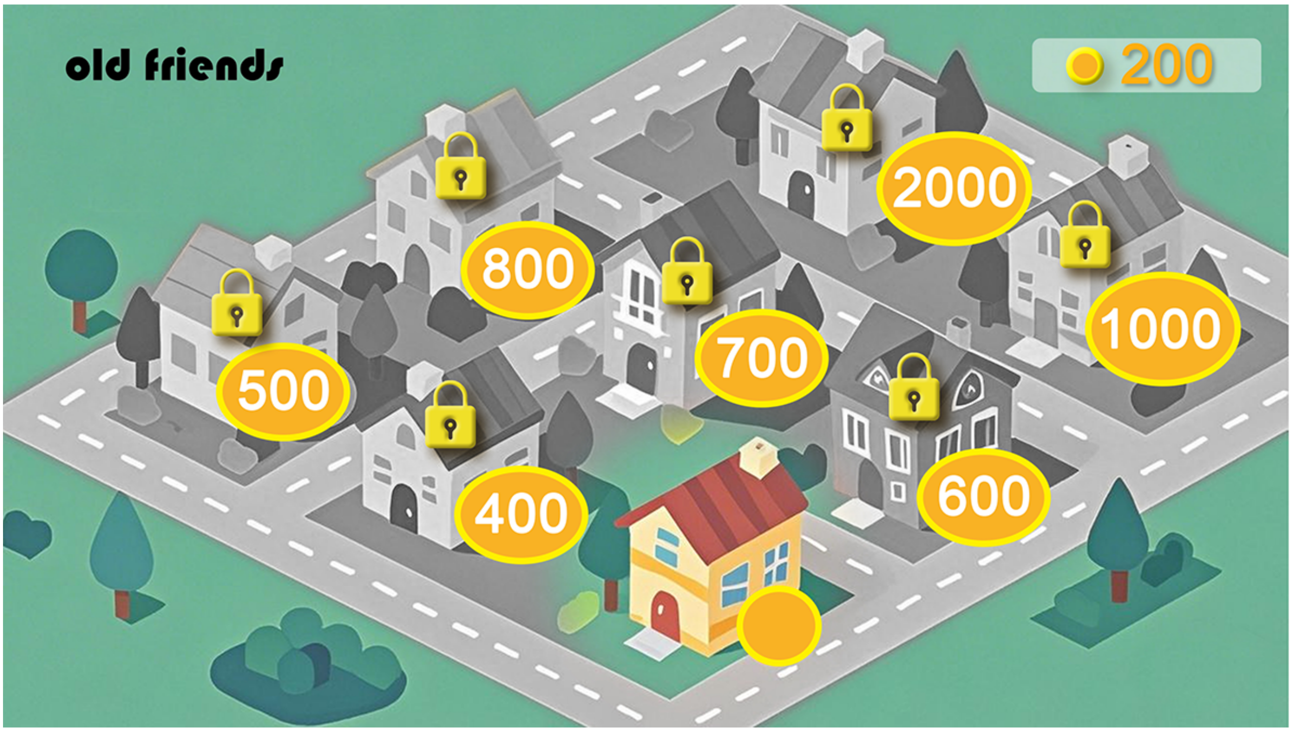


Figure 6. Failure outcome interface of the “Old Friends” serious card game.

The “Homeland” interface (Figure 7), serving as the core social and progress hub of the game, is designed to transform the abstract sense of community belonging into a concrete visual expression. The interface continues the game’s overall principle of low visual load, using soft backgrounds and high contrast numbers to ensure clear readability. Its core design feature is a series of circular blocks labeled with numbers. We chose circles because their smooth, rounded shape conveys a sense of inclusivity and harmony, perfectly aligning with the “Homeland” theme. The numbers within each block represent virtual coins needed for community building. This quantified display makes the abstract goal of “community building” concrete and controllable, thereby enhancing the clarity of tasks and operational confidence for older users. Overall, this interface is not merely a collection of functions but also a visual metaphor, as breaking down the “community” into interactive and achievable units aims to strengthen older users’ sense of participation and belonging, directly supporting the study’s core objectives of “social interaction” and “emotional support.”

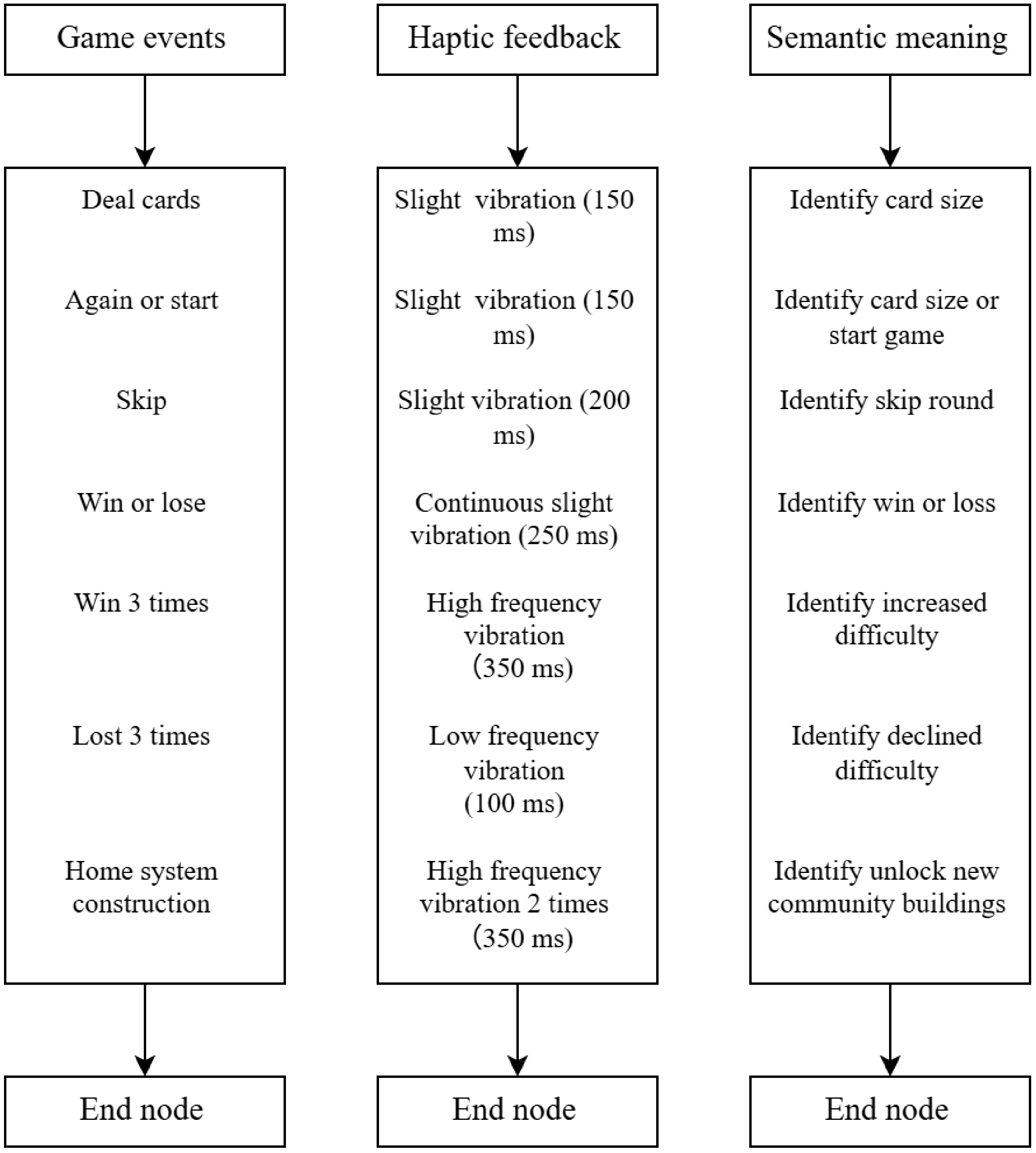
Figure 7. "Homeland" interface within the "Old Friends" serious card game.



Haptic Interaction Flowchart

To create an accessible and semantically rich gaming experience for older users, we designed a structured haptic feedback system that maps specific vibration patterns to in-game events, as shown in Figure 8. This design follows a core principle of using noninvasive haptic cues to convey game status information, thereby reducing reliance on vision and providing support for users with visual impairments.

Figure 8. Mapping of game events to structured haptic feedback patterns in the "Old Friends" serious card game.



The mapping relationship is carefully designed to ensure intuitiveness. Slight vibration (150 ms) is used for "Deal cards" and pressing the "Again" and "Start" buttons. Conversely, a continuous slight vibration (250 ms) is dedicated to important game outcomes (win or loss), providing a more significant and emotionally resonant cue. Furthermore, the system dynamically adjusts the game difficulty based on player performance, conveying this through differentiated vibrations, with a high-frequency vibration (350 ms) after 3 consecutive wins indicating an increased challenge, while a low-frequency vibration (100 ms) after 3 consecutive losses indicates a decreased difficulty. Unlocking new community buildings is achieved with 2 high-frequency vibrations (350 ms each). This closed-loop design aims to maintain an appropriate level of challenge and alleviate player frustration.

In terms of technical implementation, this haptic solution uses an Android linear resonant motor, chosen for its ability to accurately and stably generate predetermined vibration parameters. Therefore, this structured approach ensures that haptic feedback is not merely decorative but serves as an indispensable and meaningful communication channel within the game's accessibility design framework.

Discussion

Overview of Findings in Relation to Objectives

This pilot study successfully achieved the 2 main objectives outlined in the introduction. First, it systematically applied the DPE framework to design a haptic-driven prototype of serious game theory for older adults. Second, it conducted a preliminary assessment of the prototype's usability, accessibility, and user experience in the target population. The results confirmed the feasibility and preliminary acceptability of the approach. Quantitative analysis showed that the prototype achieved an excellent SUS score. In the qualitative analysis, thematic analysis revealed three key user experience themes: (1) the intuitiveness of haptic feedback, (2) its effectiveness in reducing perceived eye strain, and (3) the appeal of simplified game mechanics. These results collectively provide fundamental validation that DPE-guided design with haptic feedback as a core component can create an engaging and visually low-load cognitive training tool for older adults, thus directly achieving the objectives of this study.

Comparisons to Existing Literature

This study addresses a specific gap in the existing literature. While previous studies have successfully integrated haptic technology into serious games, their main focus has often been on enhancing realism, immersion, or motor skill training [17,19]. In contrast, our app is inherently accessibility-driven. We repurpose a common hardware feature of smartphones (vibration motors) to address a key and often overlooked problem of the pervasive visual dependence in tools designed for older populations [14,15]. Furthermore, by adopting the DPE framework, we provide a systematic, theory-based design approach that goes beyond the ad hoc or technology-centric design approaches seen in some previous work [18,20], thus providing a replicable model for future research and development in the field of inclusive aging technologies.

Interpretations

Positive user feedback regarding the intuitiveness of haptic feedback, its reduction of eye strain, and the game's ease of use validated our design approach in several ways. First, users quickly grasped the ability to identify cards through vibration cues, demonstrating that the carefully designed haptic symbol system successfully served as an effective alternative to traditional information channels. This directly confirms the feasibility of "haptic feedback as a substitute for vision" as a practical design principle [17,19]. Second, user feedback on reduced eye strain indicates that the strategy of combining high-contrast visuals with haptic cues successfully alleviated the burden on the visual system [15,32]. Finally, user appreciation for the simplified card mechanics suggests that for our target user group, reducing the cognitive burden associated with complex rules is just as important as sensory adaptation [33], ensuring user engagement and minimizing frustration.

Implications

These findings are significant. Theoretically, they provide concrete empirical support for applying sensory substitution [34] and cognitive load theories [35] to the design of technology for older adults. Our research shows that information can be strategically redistributed across different sensory channels to create a more accessible and less burdensome user experience for visually impaired older adults. Practically, the results validate the DPE framework as a feasible blueprint for building such designs. The "design" dimension ensures the learnability of haptic symbols, the "play" dimension uses these symbols to implement core mechanisms, such as difficulty adjustment, and the "experience" dimension uses them to provide emotional feedback. This provides developers with a clear path to create serious games that are not only cognitively stimulating but also inherently accessible, going beyond the practice of adding accessibility features after the fact.

Limitations

As a pilot study, this research has several limitations that must be explained to better understand the findings and guide future research. First, method limitations are significant, as the small sample size (n=10) and recruitment from a single old-age university solely through convenience sampling limit statistical power and the general applicability of the results. Second, reliance on self-reported measurements (eg, perceptions of eye fatigue during interviews) and the lack of objective physiological or cognitive indicators mean that while the reported benefits are encouraging, they remain preliminary and subjective. Third, the nature of the intervention itself also presents limitations, as a cross-sectional study design cannot assess long-term engagement, cognitive maintenance effects, or changes in the potential haptic learning curve over time. Furthermore, the technical parameters of the haptic feedback (eg, vibration pattern and duration) are based on the initial design choices and require systematic optimization.

To overcome these specific limitations, future research should prioritize the following directions: (1) conducting large-scale randomized controlled trials, including diverse populations with clinical characteristics, to verify cognitive effects using statistical methods and incorporate objective indicators; (2) using longitudinal study designs to assess users' sustained engagement, long-term adherence, and the trajectory of cognitive outcomes; (3) systematically testing haptic parameters to establish evidence-based intuitive and emotionally resonant guidelines for older adults; and (4) expanding the prototype by incorporating richer social interaction mechanisms and adaptive narratives to further enhance long-term motivation. Ultimately, transforming this framework into standardized design guidelines will be key to advancing the field of accessible cognitive training.

Conclusion

This pilot study validated a prototype of a haptic-driven serious game for older adults. The main innovation of this study lies in the systematic application of the DPE framework, putting "haptic as a substitute for vision" as a core design principle into practice. This approach fundamentally differs from existing studies, which largely use haptics to enhance overall immersion rather than addressing specific accessibility challenges posed by visual impairments. The main contribution of this study to the field is providing a reusable and theoretically sound design blueprint, offering a structured approach to creating efficient and easy-to-use cognitive training tools. Its implications in the real world, the prototype demonstrates a practical and scalable model for deploying low-visual-load interventions in communities and care facilities. These findings have broader implications, as they point to a more inclusive digital health future where therapeutic tools can be designed with sensory impairments in mind from the outset. By reducing barriers to participation, such approaches promise to promote social inclusion and support cognitive health among older adults, thereby meeting an important public health need.

Acknowledgments

We would like to thank the participants from the Dangtu Old Age University. The authors used DeepSeek and Google Translate to perform a grammar check on this article.

Funding

Funding was partly supported by a research grant from Universiti Kebangsaan Malaysia (grant GUP-2019-066).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: XH (lead), NMA (equal)
Data curation: XH (lead), NMA (equal)
Formal analysis: XH (lead), NMA (equal), YZ (equal)
Funding acquisition: NMA
Methodology: XH (lead), NMA (equal)
Resources: XH
Supervision: NMA (lead), SS (equal)
Validation: XH (lead), NMA (equal), SS (equal)
Visualization: XH
Writing–original draft: XH
Writing–review & editing: XH (lead), NMA (equal), LWL (equal)

Conflicts of Interest

None declared.

References

1. Ageing and health. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2025-11-15]
2. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56(3):303-308. [doi: 10.1001/archneur.56.3.303] [Medline: 10190820]
3. Larson EB, Yaffe K, Langa KM. New insights into the dementia epidemic. N Engl J Med 2013;369(24):2275-2277 [FREE Full text] [doi: 10.1056/NEJMp1311405] [Medline: 24283198]
4. Mouchaileh N, Hughes AJ. Pharmacological management of Parkinson's disease in older people. Pharmacy Practice and Res 2020;50(5):445-454. [doi: 10.1002/jppr.1683]
5. Del Ser T, Zea M, Valentí M, Olazarán J, López-Álvarez J, Rebollo-Vázquez A, et al. Effects of commonly prescribed drugs on cognition and mild cognitive impairment in healthy elderly people. J Psychopharmacol 2019;33(8):965-974. [doi: 10.1177/0269881119857206] [Medline: 31241413]
6. Petersen RC, Bennett D. Mild cognitive impairment: is it Alzheimer's disease or not? J Alzheimers Dis 2005;7(3):241-245. [doi: 10.3233/jad-2005-7307] [Medline: 16006668]
7. Yu J, Lam CL, Lee TMC. White matter microstructural abnormalities in amnestic mild cognitive impairment: a meta-analysis of whole-brain and ROI-based studies. Neurosci Biobehav Rev 2017;83:405-416 [FREE Full text] [doi: 10.1016/j.neubiorev.2017.10.026] [Medline: 29092777]
8. Dementia. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2025-11-04]
9. Umphred DA, Lazaro RT, Roller M, Burton G. Neurological Rehabilitation. Amsterdam, Netherlands: Elsevier Health Sciences; 2013.
10. Parkinson disease. Merck Manual Professional Edition. URL: https://www.merckmanuals.com/professional/neurologic-disorders/movement-and-cerebellar-disorders/parkinson-disease [accessed 2025-11-15]

11. González-González CS, Toledo-Delgado PA, Muñoz-Cruz V, Torres-Carrion PV. Serious games for rehabilitation: gestural interaction in personalized gamified exercises through a recommender system. J Biomed Inform 2019;97:103266 [FREE Full text] [doi: 10.1016/j.jbi.2019.103266] [Medline: 31408732]
12. Abt CC. Serious Games. London, United Kingdom: Bloomsbury Academic; 1987.
13. Benveniste S, Jouvelot P, Péquignot R. The MINWii project: Renarcissization of patients suffering from Alzheimer's disease through video game-based music therapy. In: Yang HS, Malaka R, Hoshino J, Han JH, editors. Entertainment Computing - ICEC. Berlin, Heidelberg: Springer; 2010.
14. Huang X, Ali NM, Sahrani S. Evolution and future of serious game technology for older adults. Information 2024;15(7):385. [doi: 10.3390/info15070385]
15. Farage MA, Miller KW, Ajayi F, Hutchins D. Design principles to accommodate older adults. Glob J Health Sci 2012;4(2):2-25 [FREE Full text] [doi: 10.5539/gjhs.v4n2p2] [Medline: 22980147]
16. Konstantinidis EI, Bamparopoulos G, Bamidis PD. Moving real exergaming engines on the web: the webFitForAll case study in an active and healthy ageing living lab environment. IEEE J Biomed Health Inform 2017;21(3):859-866. [doi: 10.1109/JBHI.2016.2559787] [Medline: 28113566]
17. Deng S, Chang J, Zhang JJ. A survey of haptics in serious gaming. In: De Gloria A, editor. Games and Learning Alliance. Cham: Springer International Publishing; 2014:130-144.
18. LI Y, Yoo Y, Weill-Duflos A, Cooperstock J. Towards context-aware automatic haptic effect generation for home theatre environments. In: VRST '21: Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology. 2021 Presented at: Association for Computing Machinery; December 8 - 10, 2021; Osaka, Japan p. 1-11. [doi: 10.1145/3489849.3489887]
19. Pacheco-Barrios K, Ortega-Márquez J, Fregni F. Haptic technology: exploring its underexplored clinical applications-a systematic review. Biomedicines 2024;12(12):2802 [FREE Full text] [doi: 10.3390/biomedicines12122802] [Medline: 39767709]
20. Deng S, Kirkby JA, Chang J, Zhang JJ. Multimodality with eye tracking and haptics: a new horizon for serious games? Int J Serious Games 2014;1(4):17-34. [doi: 10.17083/ijsg.v1i4.24]
21. Silva AJR, Restivo T, Gabriel J. A serious game with a thermal haptic mouse. Int J Onl Eng 2013;9(S8):74-76. [doi: 10.3991/ijoe.v9is8.3399]
22. Ferdig RE. Handbook of Research on Effective Electronic Gaming in Education (3 Volumes). IGI Global Scientific Publishing. 2009. URL: https://www.igi-global.com/book/handbook-research-effective-electronic-gaming/459 [accessed 2025-11-15]
23. Golzar J, Noor S, Tajik O. Convenience sampling. Int J Educ Lang Stud 2022;1(2):72-77. [doi: 10.4135/9781071812082.n116]
24. Su Z, Li Y, Xie Y, Huang Z, Cheng A, Zhou X, et al. Acute and long COVID-19 symptoms and associated factors in the omicron-dominant period: a nationwide survey via the online platform Wenjuanxing in China. BMC Public Health 2024 Aug 01;24(1):2086 [FREE Full text] [doi: 10.1186/s12889-024-19510-w] [Medline: 39090598]
25. Brooke J. SUS: a 'Quick and Dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.
26. Adeoye‐Olatunde OA, Olenik NL. Research and scholarly methods: semi‐structured interviews. J Am Coll Clin Pharm 2021;4(10):1358-1367. [doi: 10.1002/jac5.1441]
27. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008;24(6):574-594. [doi: 10.1080/10447310802205776]
28. Sampaio A, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. Amsterdam Waltham, MA: Elsevier/Morgan Kaufmann; 2013.
29. Appelbaum M, Cooper H, Kline RB, Mayo-Wilson E, Nezu AM, Rao SM. Journal article reporting standards for quantitative research in psychology: the APA publications and communications board task force report. Am Psychol 2018;73(1):3-25. [doi: 10.1037/amp0000191] [Medline: 29345484]
30. Willig C, Rogers WS. The SAGE Handbook of Qualitative Research in Psychology. Thousand Oaks, CA: SAGE; 2017.
31. Luo R, Wang J, Wang Y. Undergraduate students' perceptions of using videoconferencing for EFL learning: evidence from tencent meeting application. Heliyon 2023;9(12):e22993 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e22993] [Medline: 38144280]
32. Web Content Accessibility Guidelines (WCAG) 2.1. URL: https://www.w3.org/TR/WCAG21/ [accessed 2025-11-04]
33. Bilau I, Koo B, Fu E, Chau W, Kwon E, Yang E. Visual accessibility through open shelving: impacts on cognitive load, motivation, physical activity, and user perception in older adults with mild cognitive impairment. J Aging Environ 2025 Nov 12:1-28. [doi: 10.1080/26892618.2025.2576213]
34. Bach-y-Rita P, Kercel SW. Sensory substitution and the human-machine interface. Trends Cogn Sci 2003;7(12):541-546. [doi: 10.1016/j.tics.2003.10.013] [Medline: 14643370]
35. Cognitive load theory. In: Psychology of Learning and Motivation. Cambridge, MA: Academic Press; 2011:37-76.

Abbreviations

DPE: Design-Play-Experience

JARS: Journal Article Reporting Standards
SUS: System Usability Scale

Edited by S Brini; submitted 22.Oct.2025; peer-reviewed by W Chuangchai, Z Amin; comments to author 12.Nov.2025; accepted 04.Dec.2025; published 30.Jan.2026.

Please cite as:
Huang X, Mohamad Ali N, Sahrani S, Zhang Y
A Haptic-Driven Serious Game for Cognitive Stimulation and Visual Impairment Mitigation in Older Adults Based on the Design-Play-Experience Framework: Cross-Sectional Mixed Methods Pilot Study
JMIR Serious Games 2026;14:e86290
URL: https://games.jmir.org/2026/1/e86290
doi:10.2196/86290
PMID:

Correction: Tongue Muscle Training App for Middle-Aged and Older Adults Incorporating Flow-Based Gameplay: Design and Feasibility Pilot Study

Kuan-Chu Su[1], MA; Ko-Chiu Wu[2], PhD; Kuei-Ru Chou[3], PhD; Chia-Hsu Huang[2], MDes

[1]College of Design, National Taipei University of Technology, Taipei, Taiwan

[2]Department of Interaction Design, National Taipei University of Technology, Rm.701-4, Design Building, No.1, Sec.3, Chung-hsiao E. Rd, Taipei, Taiwan

[3]School of Nursing, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Ko-Chiu Wu, PhD
Department of Interaction Design, National Taipei University of Technology, Rm.701-4, Design Building, No.1, Sec.3, Chung-hsiao E. Rd, Taipei, Taiwan

Related Article:

Correction of: https://games.jmir.org/2025/1/e53045

Abstract

(JMIR Serious Games 2026;14:e90410) doi:10.2196/90410

In "Tongue Muscle Training App for Middle-Aged and Older Adults Incorporating Flow-Based Gameplay: Design and Feasibility Pilot Study" [1], the authors made one change.

The institutional affiliation of author KCS has been changed from the following:

> *Department of Interaction Design, National Taipei University of Technology, Taipei, Taiwan*

The affiliation now reads :

> *College of Design, National Taipei University of Technology, Taipei, Taiwan*

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Su KC, Wu KC, Chou KR, Huang CH. Tongue muscle training app for middle-aged and older adults incorporating flow-based gameplay: design and feasibility pilot study. JMIR Serious Games 2025 Jan 9;13:e53045. [doi: 10.2196/53045] [Medline: 39791331]

Submitted 27.Dec.2025; this is a non–peer-reviewed article;accepted 29.Dec.2025; published 16.Jan.2026.

Please cite as:
Su KC, Wu KC, Chou KR, Huang CH
Correction: Tongue Muscle Training App for Middle-Aged and Older Adults Incorporating Flow-Based Gameplay: Design and Feasibility Pilot Study
JMIR Serious Games 2026;14:e90410
URL: https://games.jmir.org/2026/1/e90410
doi:10.2196/90410

RedMan-GreenMan: Co-Designed Pedestrian Safety Game Prototype for Children With Autism

Jason Carter Stanton[1], MPhil; Nicole L Peel[2,3], MHS, MBA, PhD; Caroline J Mills[2,3], MAL, PhD; Paul P Breen[1,3], PhD

[1]The MARCS Institute for Brain, Behaviour & Development, Western Sydney University, 160 Hawkesbury Road, Westmead, Australia

[2]School of Health Sciences, Western Sydney University, Campbelltown, Australia

[3]Translational Health Research Institute, Western Sydney University, Westmead, Australia

Corresponding Author:
Paul P Breen, PhD
The MARCS Institute for Brain, Behaviour & Development, Western Sydney University, 160 Hawkesbury Road, Westmead, Australia

Abstract

This letter presents the conceptualization, design, and technical evaluation of the RedMan-GreenMan game co-designed with carers, aimed to help children with autism spectrum disorder (hereafter autism) acquire pedestrian safety skills. While the system has been implemented and is in active use, no empirical evaluation of learning outcomes or behavioral impact has been conducted to date, and the focus of this work is on system development, functionality, and technical evaluation.

(JMIR Serious Games 2026;14:e69260) doi:10.2196/69260

KEYWORDS

autism; game; pedestrian; safety; training

Introduction

Children with autism face a higher risk of unintentional injury and death compared to their peers, as challenges with attention, motor control, and cognitive delays in some children can hinder the development of essential safety skills [1,2]. In collaboration with the Skillz4me Family Centre for Disabilities, a not-for-profit organization established to support children with moderate to profound disabilities, it was identified that while some research has been completed in the area [3-5], there are currently no pedestrian road safety programs available that adequately address the specific needs of children with autism.

A specific barrier for children with autism is the difficulty in forming clear associations between related objects [6,7]. For example, a small toy train may not be readily associated with an actual full-sized train. To address this challenge, we co-designed a simple game, RedMan-GreenMan, which importantly integrates a real pedestrian traffic signal. This approach may increase the likelihood of recognition and association by the child, and in turn may support the potential transfer of safety skills from the indoor space and gameplay to the public, real-world environment.

An additional design consideration was the integration of a performance recording system to enable staff to monitor each child's interaction with the game over time.

This paper describes the design and in situ testing of the RedMan-GreenMan system. Testing aimed to validate the technical aspects of the system; it did not include any assessment of children, pedagogical outcomes, or effectiveness of the intervention in teaching safety skills.

Methods

User Requirements

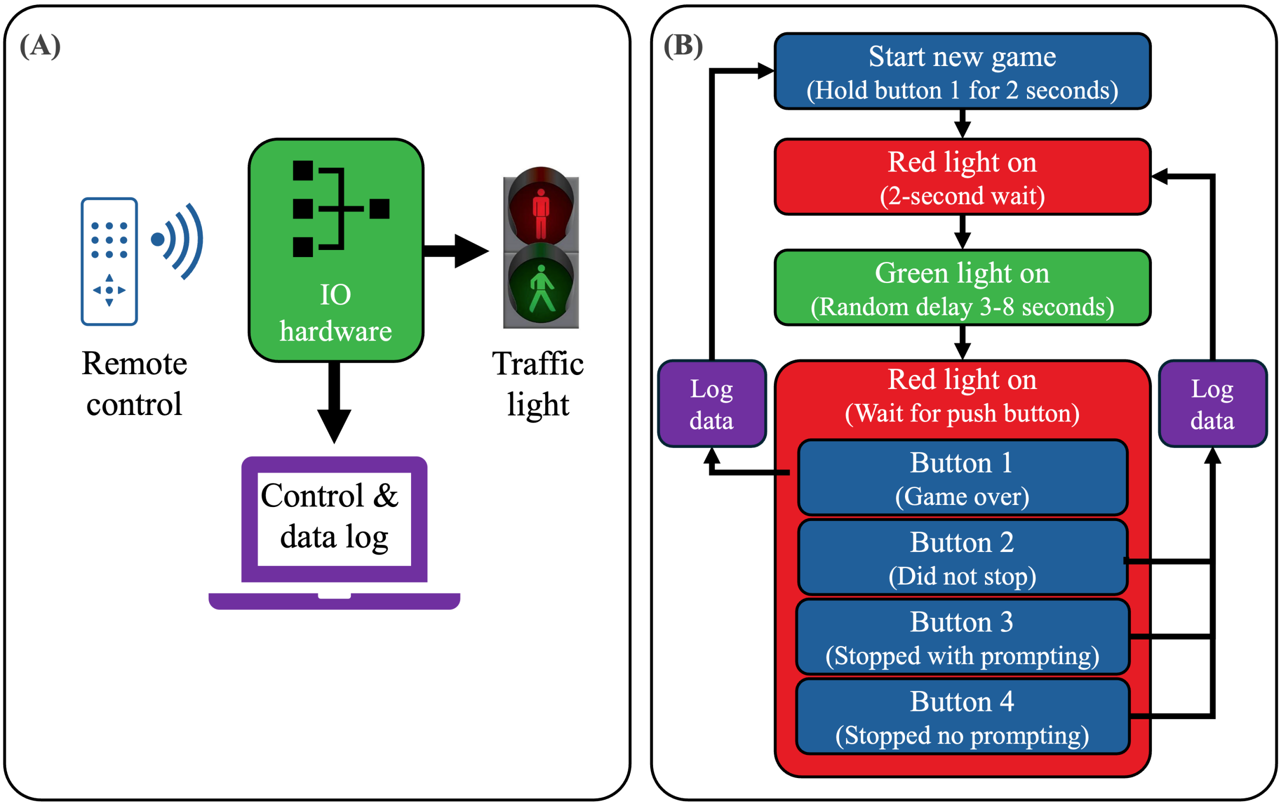
Scope for the design, development, and commissioning of the system was established. Briefly, the user requirements were:

- Use of a real pedestrian traffic light;
- Remote controllable operation of the game;
- Random timing of gameplay in operation, to prevent inadvertent signaling from staff; and
- Date-time stamping of in-game child response.

System Requirements

User requirements led to the system topology and process control as shown in Figure 1.

Figure 1. (A) System topology and (B) process control developed for RedMan-GreenMan. Key requirements included a need for the system to be controllable by a remote control to allow the staff member to interact with children within the center. The categorization of child responses was intentionally limited to three options to simplify decision-making for the busy staff member.



System Evaluation

A series of experiments was conducted to rigorously evaluate system functionality, accuracy, precision, and operational limits.

- System functionality during repeated operation cycles: With the remote positioned 5 meters from the receiver, the program was initiated by pressing and holding the "1" key for 2 seconds. A fixed button selection pattern (2-3-4-3-2) was executed and terminated by pressing the "1" key. This cycle was repeated 40 times to assess system reliability and accuracy under consistent operational conditions.
- Traffic light timing verification: Traffic light change periods were logged to a file and cross-referenced with video recordings for validation. The accuracy and precision of randomly timed light changes (range 3 - 8 s) were analyzed.
- System reliability at increasing range: System reliability was tested with 30 traffic light changes at a 2-meter distance between the remote and receiver. This process was repeated at 5-meter intervals, gradually increasing the distance until complete communication failure was observed.

A single device was used in all testing within the Skillz4me Family Centre for Disabilities with direct line of sight between the remote control and receiver up to a range of 75 m. Distances greater than 75 m lost the line of sight, as measurements were taken in the foyer (80 - 85 m), entry area (90 m), and outside the building (≥95 m).

Ethical Considerations

Ethical review was not sought as no human subjects were involved in the research.

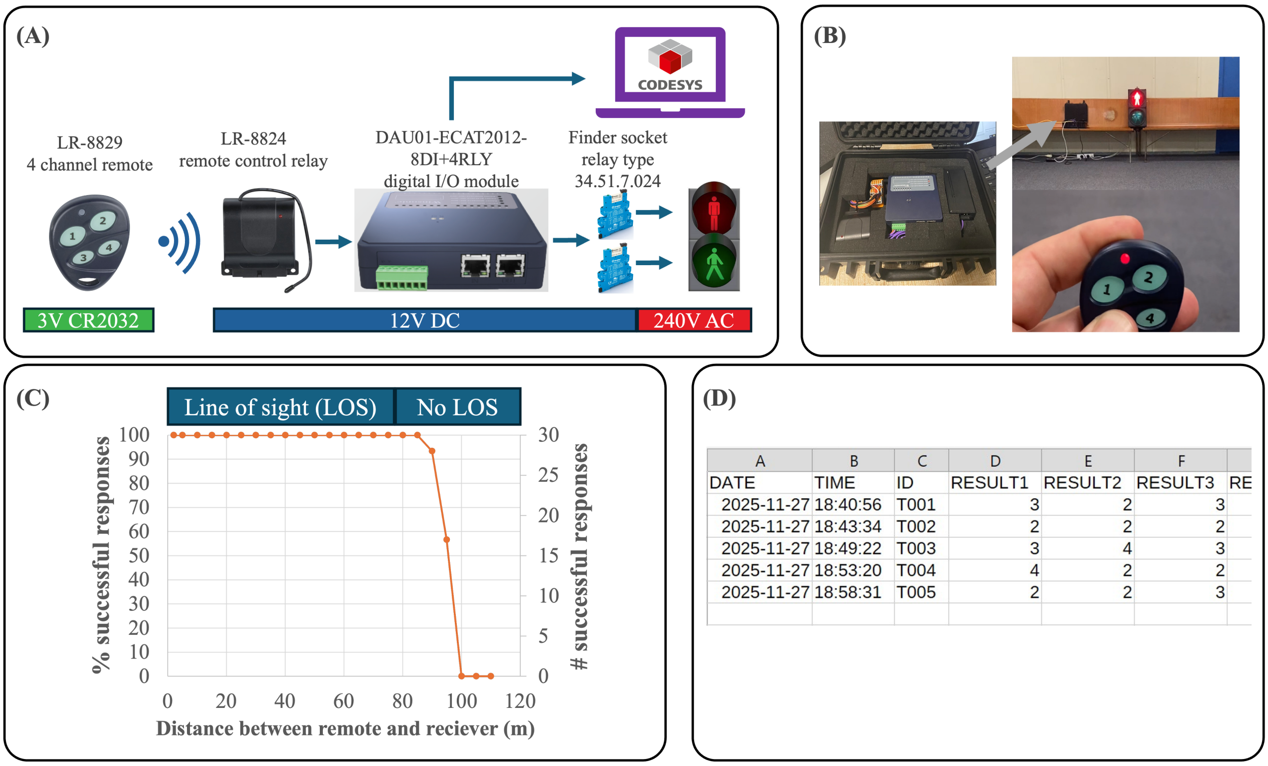
Results

An image of the developed system and its architecture is shown in Figure 2. The system's performance was evaluated through three distinct experiments. In experiment 1, a predefined sequence pattern was repeated 40 times, achieving 100% accuracy and demonstrating exceptional reliability during repeated operational cycles.

Experiment 2 focused on the accuracy of the traffic light timing, defined as the mean difference between measurements and logged reference values, which was found to be 0.096 seconds, indicating a very slight positive bias. The precision, assessed as the SD of residuals (differences between the secondary measurements and their corresponding references), was 0.072 seconds, reflecting consistent measurement performance.

Experiment 3 assessed the system's reliability over increasing distances. The system operated with 100% reliability up to a range of 85 m; this includes 10 m (from 75 to 85 m) without a direct line of sight. Beyond this range, performance declined sharply, culminating in total failure at 100 meters. It is important to note that the maximum effective range was constrained by indoor physical limitations, as direct line of sight was obstructed by the layout of the testing environment (Figure 2C). These results highlight the system's robustness and accuracy under controlled conditions while identifying potential limitations in extended-range applications.

Figure 2. (A) Developed system architecture showing specific components and power requirements. (B) Actual system created: gray arrow points to system electronics next to traffic light. (C) Reliability of system response at increasing range; in order to reach failure, the researcher had to exit the center. (D) Demonstration of a system datalog for 5 consecutive simulated games where the monitored individual (ID) is different for each game (no human subjects involved).



Discussion

This study presents the conceptualization, design, and technical evaluation of RedMan-GreenMan, a gamified intervention co-developed with the Skillz4me Family Centre for Disabilities. The system was designed to support children with autism in acquiring pedestrian safety skills by addressing challenges in associative learning. By incorporating an authentic pedestrian traffic signal, the prototype aims to bridge the perceptual and cognitive divide between the virtual learning environment and real-world traffic scenarios.

Importantly, while the system has been implemented in accordance with user-defined requirements and is currently deployed operationally at the Skillz4me Family Centre for Disabilities, no human data were collected or analyzed for this study, and no educational or pedagogical evaluation has been conducted. The current work should not be interpreted as evidence of learning impact or behavioral change in the target population.

The system's technical functionality has been validated in situ, but no empirical data is available regarding its effectiveness in teaching safety skills or supporting skill transfer to real-world contexts. Other limitations include the lack of scalability and portability of the prototype, and its construction remains cost-prohibitive (~US $350 for parts excluding laptop).

Future research should include structured educational trials to rigorously evaluate learning outcomes, skill retention, and transfer to real-world naturalistic settings. Such studies may benefit from mixed methods approaches, including behavioral observations, performance tracking, and input from caregivers or therapists. Parallel system development efforts should focus on improving affordability and portability, with the goal of enabling broader use in diverse educational and therapeutic settings.

Funding

This work received no specific grant from any funding agency.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author. Please see Multimedia Appendix 1 for system hardware design, bill of materials, and software.

Authors' Contributions

JCS led the conceptualization, data curation, original draft, and investigation. PPB led methodology, formal analysis, and visualization. PPB, CJM, and NLP equally assisted with supervision and reviewing and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
System hardware design, bill of materials, and software.
[PDF File, 668 KB - games_v14i1e69260_app1.pdf]

References

1. Lee LC, Harrington RA, Chang JJ, Connors SL. Increased risk of injury in children with developmental disabilities. Res Dev Disabil 2008;29(3):247-255. [doi: 10.1016/j.ridd.2007.05.002] [Medline: 17582739]
2. Wiseman KV, McArdell LE, Bottini SB, Gillis JM. A meta-analysis of safety skill interventions for children, adolescents, and young adults with autism spectrum disorder. Rev J Autism Dev Disord 2017 Mar;4(1):39-49. [doi: 10.1007/s40489-016-0096-7]
3. Harriage B, Blair KSC, Miltenberger R. An evaluation of a parent implemented in situ pedestrian safety skills intervention for individuals with autism. J Autism Dev Disord 2016 Jun;46(6):2017-2027. [doi: 10.1007/s10803-016-2730-8] [Medline: 26864158]
4. Dixon DR, Miyake CJ, Nohelty K, Novack MN, Granpeesheh D. Evaluation of an immersive virtual reality safety training used to teach pedestrian skills to children with autism spectrum disorder. Behav Anal Pract 2020 Sep;13(3):631-640. [doi: 10.1007/s40617-019-00401-1] [Medline: 32953391]
5. Steinborn M, Knapp TJ. Teaching an autistic child pedestrian skills. J Behav Ther Exp Psychiatry 1982 Dec;13(4):347-351. [doi: 10.1016/0005-7916(82)90083-0] [Medline: 7166609]
6. Preissler MA. Associative learning of pictures and words by low-functioning children with autism. Autism 2008 May;12(3):231-248. [doi: 10.1177/1362361307088753] [Medline: 18445733]
7. Carr D, Felce J. Teaching picture-to-object relations in picture-based requesting by children with autism: a comparison between error prevention and error correction teaching procedures. J Intellect Disabil Res 2008 Apr;52(Pt 4):309-317. [doi: 10.1111/j.1365-2788.2007.01021.x] [Medline: 18339093]

Edited by A Coristine; submitted 12.Jan.2025; peer-reviewed by E Vallefuoco, F Jorge, G Wang; revised version received 27.Nov.2025; accepted 28.Nov.2025; published 15.Jan.2026.

Please cite as:
Stanton JC, Peel NL, Mills CJ, Breen PP
RedMan-GreenMan: Co-Designed Pedestrian Safety Game Prototype for Children With Autism
JMIR Serious Games 2026;14:e69260
URL: https://games.jmir.org/2026/1/e69260
doi:10.2196/69260

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org