Protocol

Immersive Virtual Reality Exergaming vs Video-Guided Screen-Based Exercise for Cognitive and Neurophysiological Adaptations in Young Adults: Protocol for Acute and Chronic Experimental Studies

Miguel Becerra-Alonso[1]; Sonia Ortega-Gómez[2,3]; Borja Sañudo[2,4]; Adrián Feria-Madueño[2,4]

[1]Department of Education and Sport, Universidad Loyola Andalucía, Seville, Spain

[2]Department of Physical Education and Sport, University of Seville, Seville, Spain

[3]MOVE-IT Research Group, Department of Physical Education, Biomedical Research and Innovation Institute of Cádiz (INiBICA), University of Cádiz, Cádiz, Spain

[4]Asociación de Investigación e Innovación en Ejercicio Físico y Tecnologías Aplicadas al Movimiento y la Salud, Seville, Spain

Corresponding Author:
Sonia Ortega-Gómez
Department of Physical Education and Sport
University of Seville
Faculty of Education Sciences, C/ Pirotecnia s/n
Seville, 41013
Spain
Phone: 34 955420585
Email: sortega@us.es

Abstract

Background: Physical exercise involving simultaneous cognitive demands, including dance-based exergames, has shown moderate effects on executive function. Immersive virtual reality (VR) may enhance these effects by increasing presence and frontal cortical activation; however, evidence in youth integrating objective neurophysiological measures remains limited and methodologically heterogeneous.

Objective: This study aims to examine the acute and chronic effects of an immersive VR dance exergame, considered as an integrated package of immersion, interactivity, and real-time performance feedback, compared with a matched video-guided screen-based exercise condition (comparator condition [CS]), on inhibitory control as the primary outcome, and on frontal cortical activity, perceived self-efficacy, affective state, and physiological response as mechanistic or secondary outcomes.

Methods: This protocol reports 2 methodologically independent studies, covered by a single prospective trial registration that describes both, in young adults aged 18 to 30 years: an acute randomized crossover study with up to 40 participants, and a 10-week 2-arm randomized controlled trial with a target sample of 64 participants and 3 sessions per week. In the VR condition, participants will play Beat Saber using Meta Quest 2 headsets. In the CS, they will follow a prerecorded video of the same playlist, matched for songs, tempo, movement sequence, difficulty, transition periods, and session duration. The primary outcome is the Stroop interference index. The confirmatory neurophysiological end point is absolute frontal theta power (4-7 Hz) averaged across F3, F4, AF3, and AF4 during the color-word interference condition. Secondary and exploratory outcomes include additional spectral measures, perceived self-efficacy, affective state, heart rate and percentage of heart rate reserve, rating of perceived exertion, wrist accelerometry, presence, cybersickness, enjoyment, and expectations. Assessments are conducted before and after each session in the acute study and at baseline, midpoint, postintervention, and 4-week follow-up in the chronic study. Linear mixed models accounting for period, sequence, participant, and the within-session pre/post structure will be used, with a prespecified hierarchy of confirmatory and exploratory analyses and false discovery rate correction applied within defined families.

Results: Recruitment is expected to begin in October 2026, with data collection completed by December 2027 and first results available by June 2028. No participants had been enrolled at submission.

Conclusions: This protocol presents a controlled experimental design to examine the effects of immersive VR exergaming, compared with matched video-guided screen-based exercise, on neurocognitive, psychological, and physiological outcomes.

XSL•FO
RenderX

Incorporating immersion-related variables may help clarify the mechanisms underlying VR-based exercise and inform future technology-supported interventions in exercise and cognitive health.

Trial Registration: ISRCTN Registry ISRCTN15563829; https://www.isrctn.com/ISRCTN15563829

International Registered Report Identifier (IRRID): PRR1-10.2196/105801

(JMIR Res Protoc 2026;15:e105801) doi: 10.2196/105801

KEYWORDS

crossover; EEG; executive function; randomized controlled trial; self-efficacy; study protocol; virtual reality

Introduction

Scientific evidence has consistently shown that physical exercise is a key modulator of brain function, promoting neural plasticity, emotional regulation, and executive control [1,2]. In young populations, these adaptations include both acute effects, such as increased cerebral blood flow, neurotransmitter release, and activation of frontoparietal networks, and chronic adaptations related to improvements in functional connectivity and synaptic efficiency. From a mechanistic perspective, exercise increases neurotrophic factors such as brain-derived neurotrophic factor, which is involved in learning and synaptic reorganization, and contributes to the regulation of the hypothalamic-pituitary-adrenal axis, sleep quality, and emotional stability [2].

Scientific interest has progressively shifted toward intervention models that integrate physical and cognitive demands simultaneously, commonly referred to as dual-task or cognitive-motor training. Within this paradigm, active video games or exergames based on dance choreographies require rhythmic synchronization, motor memory, error monitoring, and real-time decision-making, creating a highly demanding perceptual and cognitive environment. Recent meta-analytic evidence indicates that exergames may produce moderate effects on executive function, particularly on inhibition and cognitive flexibility, although the available evidence remains methodologically heterogeneous [3].

Immersive virtual reality (VR) offers a unique combination of perceptual immersion, embodied interaction, and enhanced feelings of presence that distinguish it from traditional digital interventions [4-6]. Emerging evidence suggests that these characteristics may promote greater engagement, motivation, attentional focus, and self-regulatory behaviors, particularly in younger populations [7-10]. However, a recurrent limitation in research on exergames and VR is the lack of assessment of key variables required to interpret the mechanisms underlying VR-related effects, such as presence, cybersickness, enjoyment, motivation, and participant expectations. Without these measures, it is difficult to determine whether differences between VR and conventional screen conditions are attributable to the combined immersive and interactive characteristics of VR or to other factors, such as device novelty or participant expectations. Recent reviews of immersion and presence in head-mounted display environments have refined these constructs and their measurement [11-13].

Another frequent limitation is the reliance on behavioral measures alone, without incorporating objective neurophysiological indicators. Electroencephalography can address this limitation by allowing the direct assessment of cortical activity with high temporal resolution, particularly in the frontal theta, alpha, and beta bands, which are related to mental workload, attentional regulation, and executive control [14-16]. Evidence in VR and interactive dance contexts in young adults remains limited, with few studies integrating neurophysiological and behavioral measures under controlled experimental conditions [17-19].

Two randomized trials are particularly relevant to the present question and delimit the knowledge gap it addresses. Huang [20] compared immersive and nonimmersive exergaming over 4 weeks in adults older than 50 years, a population in which age-related decline in executive function is expected, and reported greater improvement in selected executive-function outcomes under the immersive condition. Lachowicz et al [21] compared immersive VR training using the same commercial exergame against a no-training control condition in amateur esports athletes and reported superior outcomes in the trained groups. Together, these trials suggest that immersive exergaming can improve selected executive-function outcomes relative to no training, and that immersive delivery may outperform nonimmersive delivery in an older population. However, several methodological and mechanistic questions remain.

This protocol addresses these gaps in 4 ways. First, the comparator is an active condition matched for music, song order, tempo, movement sequence, difficulty, session duration, and transition periods, which reduces the interpretational ambiguity associated with a no-training control. Second, acute and chronic responses are examined within a single protocol using the same task and outcome battery, allowing single-session responses and responses to repeated exposure to be compared on a common measurement scale. Third, behavioral and neurophysiological outcomes are integrated through a prespecified brain-behavior link. Fourth, presence, cybersickness, enjoyment, and prior expectations are assessed as prespecified potential mediators and moderators, allowing experiential mechanisms to be examined directly. The population also differs from both trials, since participants are nonathletic young adults rather than older adults or trained esports players.

The rationale for studying cognitively healthy young adults is mechanistic rather than clinical. The aim is not to remediate a deficit or to enhance executive function beyond the normal range, but to determine whether the delivery medium modulates the acute and adaptive cognitive response to a standardized cognitive-motor task, and if so, through which pathway.

XSL•FO
RenderX

Answering this question requires a sample in which the outcome is not constrained by pathology or age-related decline, because in older or clinical populations any between-condition difference may be confounded by differential baseline impairment, capacity to engage with the technology, or room for recovery. Young adults therefore serve as a model population in which the effect of the medium can be examined with fewer confounding influences. Establishing whether the medium matters, and through which mechanism, is a necessary step before designing trials in populations in which such effects may have direct clinical relevance. Transient modulation of inhibitory control within the normal range may also be relevant in educational and training contexts, where the question is whether a short bout of activity improves the cognitive state in which subsequent learning occurs.

Overall, important gaps remain regarding experimental control, task standardization, objective assessment of brain activation, and the evaluation of variables that may mediate or moderate the effect of the delivery medium. Addressing these gaps is necessary to determine whether the potential benefits of VR-based exercise are explained by the properties of the immersive interactive medium, by differences in motor or physiological load, or by psychological factors such as presence, enjoyment, and expectations. Therefore, this paper describes the protocol of 2 independent studies designed to examine the acute and chronic effects of an immersive VR dance exergame compared with a matched video-guided screen-based exercise condition, using the same songs, tempo, movement sequences, and difficulty level, verifying the equivalence of physiological and motor load, and evaluating potential neurocognitive mechanisms, in accordance with the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2025 recommendations for reporting clinical trial protocols [22].

We hypothesize that, compared with the matched video-guided screen-based condition, the immersive VR condition will produce greater frontal theta power and greater improvement in inhibitory control, as well as higher perceived self-efficacy and enjoyment, while maintaining comparable physiological and motor load. For the chronic study, we specifically adopt the cumulative-effect hypothesis: we predict that the cognitive and neurophysiological response to repeated exposure accumulates across sessions, such that the between-condition difference at week 10 exceeds that observed at the midpoint assessment. The follow-up assessment will be used exploratorily to examine whether any postintervention differences are maintained after the end of the program. Because the 2 conditions differ concurrently in immersion, interactivity, and real-time performance feedback, these hypotheses concern the immersive exergame as an integrated package and are not hypotheses about immersion in isolation.

Methods

Study Design

Overview

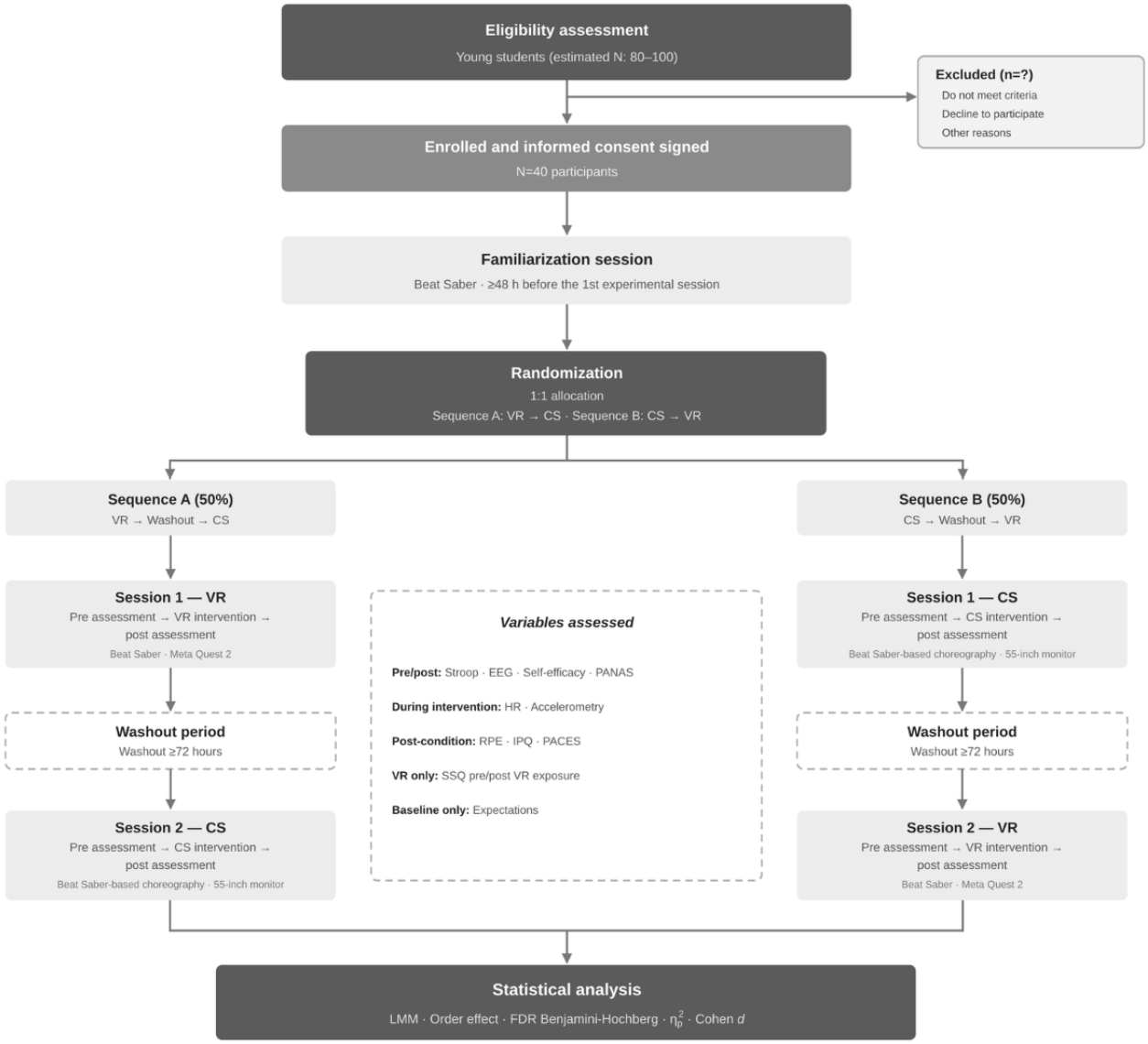
This paper reports the protocols of 2 methodologically independent studies, developed in accordance with the SPIRIT 2025 recommendations. The 2 studies recruit independent, nonoverlapping samples and differ in design, allocation procedure, sample size, exposure schedule, assessment time points, and analysis plan. Both studies are covered by a single prospective registration in the International Standard Randomised Controlled Trial Number (ISRCTN), the record of which has been extended to describe each study separately, including its design, allocation procedure, target sample size, exposure schedule, outcome time points, and analysis plan, so that each is identifiable and prospectively specified within the record. The description of the intervention follows the TIDieR (Template for Intervention Description and Replication) checklist [23]. Completed SPIRIT and TIDieR checklists, the prespecified statistical analysis plan, the GAIDeT (Generative AI Delegation Taxonomy) declaration, and the participant information sheet and informed consent form are provided as Multimedia Appendices 1-5.

Acute Phase

A randomized crossover design will be used, in which each participant will complete both experimental conditions, immersive VR and the matched video-guided screen-based condition (comparator condition [CS]), in a counterbalanced order. Participants will be allocated in a 1:1 ratio to one of two possible sequences: VR followed by CS, or CS followed by VR. Each experimental session will include a preassessment, the assigned intervention condition, and a postassessment, so that each participant contributes 2 measurements within each of the 2 periods. A minimum washout period of 72 hours will be established between sessions. The rationale for the crossover design is that each participant serves as their own control, which removes between-participant variance from the comparison and increases precision for a given sample size. The design does not by itself control carryover effects; carryover is addressed through the washout period, the counterbalancing of order, and the prespecified statistical evaluation of period, sequence, and residual effects described in the Statistical Analysis section. The washout period of at least 72 hours is justified on the basis that the transient modulation of executive function following a single bout of exercise is largest immediately after exercise and diminishes over the subsequent minutes to hours, being attributed to transient changes in arousal, cerebral perfusion, and catecholaminergic activity, so that an interval of 72 hours is several orders of magnitude longer than the documented duration of this response [24]. This interval is deliberately not expected to erase training adaptation, which is a distinct process operating over weeks and which is the object of the chronic study. Participants in the acute study complete 2 exposures in total, which does not constitute a training stimulus of the kind hypothesized to produce cumulative adaptation. Figure 1 presents the flow diagram of the acute crossover study.

XSL•FO
RenderX

Figure 1. Flow diagram of the acute randomized crossover phase. Participants will be allocated in a 1:1 ratio to one of two counterbalanced sequences: virtual reality (VR) followed by video-guided screen-based exercise on a conventional screen (comparator condition [CS]), or CS followed by VR. Each experimental session will include a preassessment, the assigned intervention condition, and a postassessment, with a washout period of at least 72 hours between sessions. EEG: electroencephalography; HR: heart rate; IPQ: Igroup Presence Questionnaire; PACES: Physical Activity Enjoyment Scale; RPE: rating of perceived exertion; SSQ: Simulator Sickness Questionnaire.

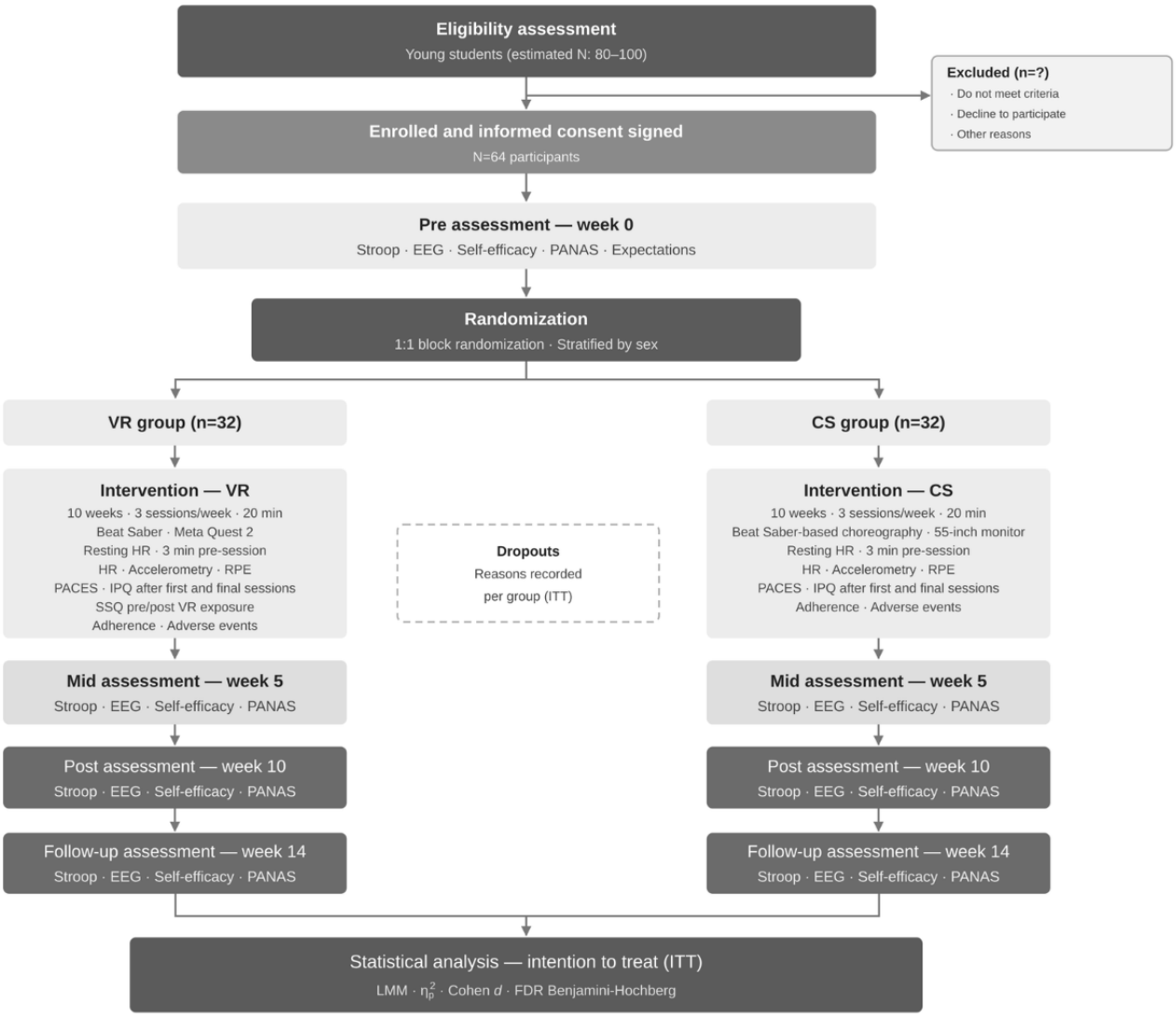


Chronic Phase

A 2-arm randomized controlled trial (RCT) will be conducted, with participants allocated to either the VR or the CS for 10 weeks. The intervention will include 3 sessions per week, resulting in a total of 30 sessions. The chronic study primarily tests the cumulative-effect hypothesis, namely that the response to repeated exposure accumulates across sessions. Because accumulation cannot be evaluated with two measurement points, assessments will be conducted at baseline (week 0), at the midpoint of the intervention (week 5), immediately after the intervention (week 10), and at a 4-week follow-up (week 14). The week 5 assessment provides an intermediate point to characterize the trajectory of change during the intervention, whereas the week 14 assessment is retained as a secondary exploratory follow-up to examine the persistence of any changes after the end of the program. Figure 2 presents the flow diagram of the chronic RCT. Both flow diagrams were developed in accordance with the CONSORT (Consolidated Standards of Reporting Trials) recommendations [25].

XSL•FO
RenderX

Figure 2. Flow diagram of the chronic randomized controlled trial phase. Participants will be allocated 1:1 to the virtual reality (VR) or video-guided screen-based exercise on a conventional screen (comparator condition [CS]) group and will complete a 10-week intervention with three 20-minute sessions per week. Main assessments will be conducted at weeks 0, 5, 10, and 14, with session-level monitoring throughout the intervention. EEG: electroencephalography; HR: heart rate; IPQ: Igroup Presence Questionnaire; PACES: Physical Activity Enjoyment Scale; RPE: rating of perceived exertion; SSQ: Simulator Sickness Questionnaire.



Participants, Recruitment, and Study Timeline

The project is structured around 2 complementary yet methodologically independent studies designed to examine both the acute and chronic effects. The chronic intervention study will recruit a target sample of 64 participants drawn from higher vocational training programs and university degree courses, allowing the evaluation of longer-term adaptations following repeated exposure to immersive VR. In parallel, an independent sample of up to 40 participants will be recruited for an acute crossover study aimed at investigating the immediate cognitive and neurophysiological responses elicited by a VR session under controlled experimental conditions. All assessments and interventions will be conducted at the same research facility using standardized procedures and equipment. Given the growing evidence of sex-related differences in executive functioning, cortical activity, sensory processing, and responses to immersive virtual environments, sex will be considered a potential confounding factor throughout the study. In the chronic RCT, sex-stratified block randomization will be used to ensure balanced group allocation. In the acute crossover phase, where each participant serves as their own control, sex will be balanced across the 2 intervention sequences and considered as a potential covariate in the statistical analyses. This approach is expected to enhance the internal validity of the study and improve the interpretability of both acute and chronic intervention outcomes. The inclusion criteria will be age between 18 and 30 years; absence of neurological, cardiovascular, or musculoskeletal conditions incompatible with moderate-intensity physical exercise; sufficient functional capacity to perform the activity; and availability to complete the procedures of the assigned study phase.

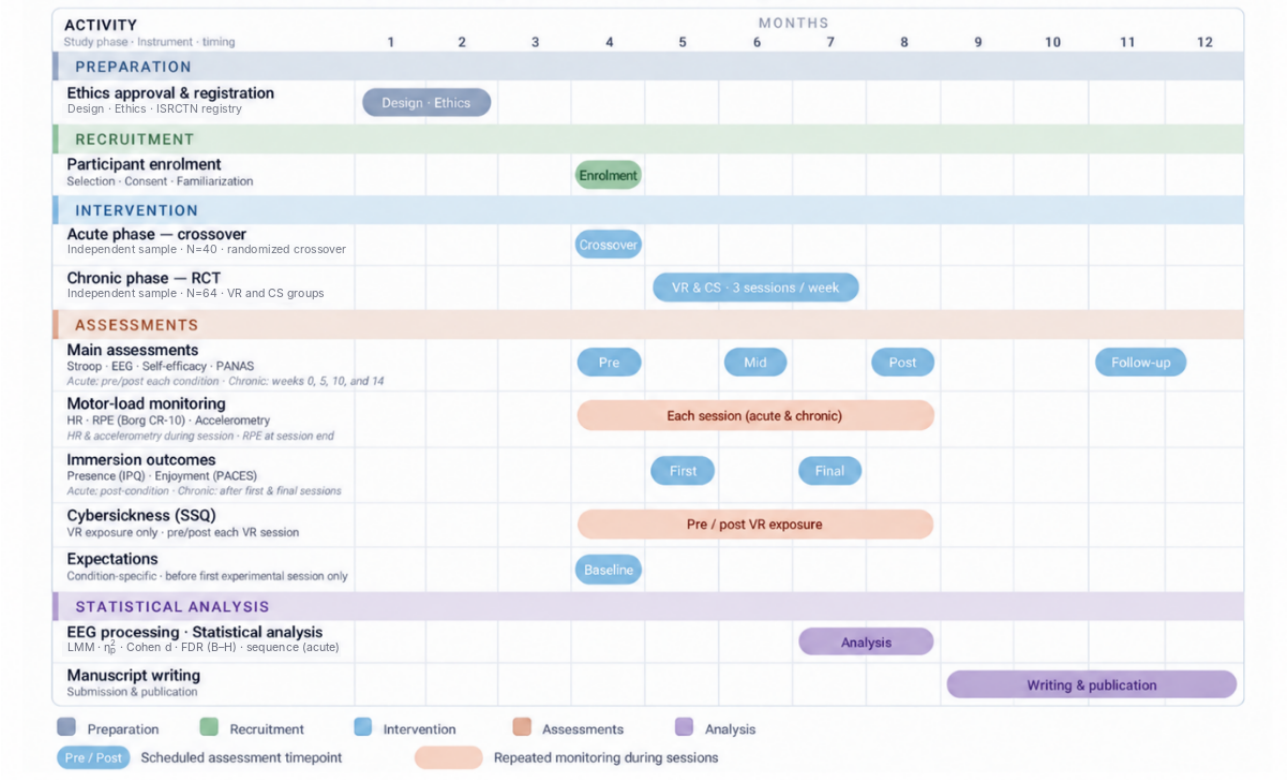
Participants will be excluded if they have a diagnosis of neurological disorders, including photosensitive epilepsy, due

XSL•FO
RenderX

to the rapid visual stimulation inherent to the video game. Participants with vestibular disorders or a history of severe dizziness will also be excluded because of the increased risk of cybersickness associated with the use of VR headsets. Current use of psychotropic medication will constitute another exclusion criterion, given its potential confounding effect on the cognitive and neurophysiological outcomes assessed. Participants with uncorrected visual impairments or difficulties in color perception, such as color blindness, will also be excluded because of their possible interference with the Stroop Color and Word Test. Previous experience with immersive VR will be assessed at baseline. Participants reporting regular engagement with Beat Saber or similar rhythm-based VR exergames (≥1 session/week during the previous 6 months) will be excluded to minimize task-specific learning effects and potential ceiling effects. General prior experience with immersive VR technology will not constitute an exclusion criterion; instead, it will be recorded and considered as a potential covariate or moderator in sensitivity analyses. This approach is consistent with current recommendations emphasizing the assessment and reporting of prior VR familiarity while avoiding arbitrary exposure thresholds that may unnecessarily restrict external validity [8,26].

The general study timeline is presented in Figure 3, developed in accordance with SPIRIT recommendations. This figure summarizes the main stages of the protocol, including recruitment, informed consent, familiarization, pre- and postassessments, the intervention period, and the recording of adherence and adverse events.

Figure 3. SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2025 schedule of enrollment, interventions, and assessments. The acute crossover and chronic randomized controlled trial (RCT) phases are conducted in independent samples. “Pre/Post” and “Baseline” indicate scheduled assessment time points: in the acute sample, assessments are conducted pre/post each condition, with expectations assessed at baseline only; in the chronic sample, baseline, midpoint, postintervention, and follow-up assessments are conducted at weeks 0, 5, 10, and 14, respectively. Immersion measures (Igroup Presence Questionnaire [IPQ] and Physical Activity Enjoyment Scale [PACES]) are assessed after the first and final intervention sessions. Heart rate (HR), rating of perceived exertion (RPE), and accelerometry are recorded each session; Simulator Sickness Questionnaire (SSQ) is administered before/after virtual reality (VR) exposure only. FDR: false discovery rate; PANAS: Positive and Negative Affect Schedule.



Sample Size Calculation

The sample size calculation was performed a priori using G*Power software (version 3.1.9.7) for the chronic phase, considering the Stroop interference index as the primary outcome. The assumed effect size (f=0.25) was based on previous meta-analytic evidence reporting small to moderate effects of physical-cognitive interventions and exergames on executive function outcomes, particularly inhibitory control and cognitive flexibility. Specifically, Stanmore et al [3] reported overall small to moderate improvements in cognitive performance following exergame interventions, while Ludyga et al [27] observed acute exercise-related benefits on executive function within a similar magnitude range. Given the methodological heterogeneity of the available literature and the absence of directly comparable studies combining immersive VR, dance-based exergaming, and electroencephalography outcomes, a conservative moderate effect was selected to avoid overestimating the expected intervention effect. Accordingly, the sample size calculation assumed f=0.25, α=.05, and 80% statistical power [3,27]. The selected model was a mixed repeated-measures ANOVA with 2 groups and 2 time points. Under these assumptions, the minimum required sample size was 54 participants. Considering an anticipated dropout rate of

XSL•FO
RenderX

15% to 20%, the target sample size was set at 64 participants, equally allocated to both conditions. For the acute crossover phase, an a priori sample size calculation was also performed using G*Power software (version 3.1.9.7), considering the Stroop interference index as the primary outcome. The calculation was based on a repeated-measures within-subject design with 2 conditions, assuming a moderate effect size (f=0.25), α=.05, 80% statistical power, 2 repeated measurements, a correlation among repeated measures of 0.50, and a nonsphericity correction of 1. Under these assumptions, the minimum required sample size was 34 participants. To account for potential dropouts, missing data, or invalid electroencephalography recordings, the target sample size for the acute crossover phase was set at 40 participants. Although electroencephalography-derived variables constitute key mechanistic outcomes, the study was not specifically powered to detect changes in neurophysiological measures. Therefore, electroencephalography findings will primarily contribute to the interpretation of potential mechanisms underlying behavioral changes observed in executive function. The week 5 midpoint assessment and the week 14 follow-up assessment are additional repeated measurements intended to characterize the trajectory and maintenance of change; they do not change the primary postintervention comparison on which the sample size calculation was based. The studies are powered for the behavioral primary outcome and not for the neurophysiological outcomes, which is why the latter are designated as mechanistic rather than coprimary, and a smallest effect of interest is prespecified for the confirmatory electroencephalographic end point in the Statistical Analysis section.

Intervention

Video Game and Hardware

The intervention will be based on the active dance exergame *Beat Saber* (Reality Labs/Meta, 2019), selected because of its guided choreography structure with synchronized visual and auditory stimuli and its use in research contexts. The task requires visuomotor coordination, rhythmic synchronization, motor memory, and real-time decision-making.

The exercise task will consist of a standardized playlist from *Beat Saber* Original Soundtrack Vol 1, performed at the easy difficulty level with the "No Fail" mode activated to avoid interruptions due to performance errors. The same playlist, order of songs, difficulty level, rhythm, movement sequence, and transition periods will be used in both experimental conditions. The playlist will include 9 songs: "$100 Bills" (2:21; 210 beats per minute [BPM]), "Beat Saber" (1:50; 166 BPM), "Breezer" (2:01; 112 BPM), "Commercial Pumping" (2:21; 105 BPM), "Country Rounds–Sqeepo Remix" (2:50; 210 BPM), "Escape ft Summer Haze" (2:41; 175 BPM), "Legend ft Backchat" (1:57; 90 BPM), "Lvl Insane" (1:50; 160 BPM), and "Turn Me On ft Tiny C" (1:57; 80 BPM). The total active gameplay time will be 19 minutes 48 seconds. The tempo of the selected songs will range from 80 to 210 beats per minute, according to the information displayed in the game.

A standardized transition period of approximately 25 seconds will be included between consecutive songs, corresponding to the time required to move from one song to the next in the game. The prerecorded video used in the CS will include the same transition periods between songs. Therefore, the active exercise component will be equivalent across conditions, and the total elapsed task duration will also be standardized. Active gameplay time is therefore 19 minutes 48 seconds, and total task duration, including the 8 intersong transitions, is approximately 23 minutes; both values are reported here to avoid the ambiguity of describing the session by a single rounded figure.

The first song will be used as an initial familiarization segment within the session, allowing participants to adapt to the rhythm, visual cues, and required movement patterns. The following songs will constitute the main exercise segment. The final song, which has the lowest tempo in the playlist, will provide a lower-tempo ending segment.

VR Condition

Participants will use Meta Quest 2 headsets, with a 120° field of view, 90 Hz refresh rate, 1832 × 1920 px resolution per eye, and a 6-degrees-of-freedom optical tracking system. Participants will use wireless controllers to interact with the 3D virtual environment. In this condition, participants will perform the selected *Beat Saber* playlist directly in the immersive virtual environment, receiving real-time task-related feedback, score progression, and visual cues inherent to the game.

Comparator Condition

The CS is a video-guided screen-based exercise delivered on a conventional screen and is designated CS throughout this paper. Participants will use a 55-inch LED monitor, with a resolution of 3840 × 2160 px and a refresh rate of 60 Hz, positioned frontally at a distance of 1.5 meters. Participants will watch a prerecorded video of a researcher performing the same Beat Saber playlist, using the same songs, order, difficulty level, tempo, movement sequence, and transition periods as in the VR condition, and will reproduce the observed movements using foam sticks that simulate the wireless controllers used in the VR condition. Audio will be provided through stereo speakers calibrated in both conditions, with an equivalent sound pressure level of approximately 70 dB. This condition is explicitly not an interactive exergame: participants do not interact with the game, do not receive a score, and do not receive real-time performance feedback. It is a matched exercise comparator in which the motor and musical content of the task is preserved, and it corresponds to the form in which home-based exercise is most commonly delivered.

Both conditions are therefore matched in active gameplay duration, song order, musical structure, tempo, difficulty level, transition periods, and expected motor pattern, and differ concurrently in immersion, interactivity, real-time performance feedback and goal structure. These features are manipulated as a package and cannot be separated by this design, which is stated explicitly in the Limitations. No 2D version of the task used exists, so an interactive screen-based comparator would require a different game and would break the matching of music, choreography and movement pattern that the present comparator achieves; the comparator was therefore chosen to preserve motor and musical equivalence, and the claims made are restricted accordingly. The difficulty level will remain constant throughout

XSL•FO
RenderX

the study, and no progression in difficulty will be introduced, in order to minimize the influence of motor learning when comparing conditions.

All participants will complete a familiarization session before the start of the study, at least 48 hours before the first experimental session, to reduce learning effects related to the device and the task. During this session, a 3-minute resting-state electroencephalography recording will additionally be obtained using the same procedures as in the experimental sessions, to provide the first of the two measurements from which the test-retest reliability of the confirmatory end point is estimated. No additional motivational reinforcement, rewards, or researcher-delivered feedback will be provided in either condition, thereby ensuring that both groups receive equivalent levels of investigator interaction throughout the study period. Fidelity of delivery will be monitored with a session checklist recording completion of the full playlist, any interruption, any deviation from the standardized script, and the researcher's presence. No tailoring or modification of the intervention is permitted.

Control of Motor Load Equivalence

A critical assumption of the study design is that physiological and motor load should be comparable between conditions [28], because a difference in exercise intensity would provide an alternative explanation for any difference in cognitive outcomes. Continuous heart rate (HR) will be recorded in every session using a Polar H10 chest strap monitor [29,30], preceded by a 3-minute seated baseline recording. Mean HR, estimated maximum HR, and percentage of HR reserve (%HRR) computed with the Karvonen formula will be derived for every session. Maximal HR will be estimated using the equation of Tanaka et al [31], $208 - 0.7 \times \text{age}$. Triaxial accelerometry will be recorded using a wrist-worn ActiGraph wGT3X-BT accelerometer at 100 Hz on the dominant wrist. Estimated energy expenditure derived from the wrist accelerometry signal will be reported as a complementary objective indicator of external load. Rating of perceived exertion (RPE) will be assessed at the end of each session using the Borg CR-10 scale [32] and is retained as a subjective complement to these objective measures rather than as the means of matching intensity. Equivalence of intensity between conditions will be evaluated by equivalence testing rather than by the absence of a significant difference, as specified in the Statistical Analysis section, with %HRR as the principal intensity variable and an equivalence margin of 5 percentage points of HR reserve. If equivalence is not established, the intensity variable concerned will be entered as a covariate in the primary model, a sensitivity analysis will be conducted with and without that adjustment, and the interpretation of any between-condition difference will be qualified explicitly.

Outcome Measures and Instruments

Overview

The primary outcome of both studies is the Stroop interference index derived from the Stroop Color and Word Test. This outcome was selected because inhibitory control is the executive function most consistently associated with acute and chronic exercise-related cognitive adaptations, and it served as the basis for the a priori sample size calculation [27]. The single confirmatory neurophysiological end point is absolute theta power in the 4 to 7 Hz band, averaged across the 4 frontal electrodes F3, F4, AF3, and AF4 as a single region rather than analyzed electrode by electrode, recorded during the color-word interference condition of the Stroop task; the comparison constituting the hypothesis is the condition-by-time interaction in the model specified in the Statistical Analysis section. Frontal theta activity has been consistently associated with cognitive control, attentional allocation, conflict monitoring and inhibitory processing, and changes in this end point will be interpreted as a candidate neurophysiological mechanism underlying observed changes in executive performance [33]. Testing proceeds hierarchically: the behavioral primary outcome is tested first, and the confirmatory neurophysiological end point second. All remaining spectral measures, and all psychological, physiological and immersion-related variables, are designated secondary or exploratory, as specified in Table 1, and no additional outcome will be promoted to confirmatory status after the data have been inspected.

XSL•FO
RenderX

Table 1. Variables, instruments, assessment time points, outcome classification, score range and planned analyses.

Variable	Instrument	Assessment time point	Outcome classification	Score range and interpretation	Planned analysis
Inhibitory control (Stroop interference index)	Stroop Color and Word Test, sheet-based format	Pre/post each condition (acute phase); weeks 0, 5, 10, and 14 (chronic)	Primary	Difference score (CW[a] − CW′[b]); higher=better interference control	LMM[c]: condition/group × time; period and sequence included in the acute crossover phase
Frontal theta power (F3, F4, AF3, AF4)	Emotiv EPOC X EEG[d] system, 256 Hz	Resting state and Stroop task pre/post each condition (acute); weeks 0, 5, 10, and 14 (chronic)	Confirmatory (mechanistic)	Continuous; smallest effect of interest and ICC[e] prespecified	LMM: condition/group × time; period and sequence included in the acute crossover phase
Perceived self-efficacy (global score)	Adapted Bandura-based self-efficacy scale	Pre/post each condition (acute); weeks 0, 5, 10, and 14 (chronic)	Secondary	0-100; higher=greater self-efficacy	LMM
Positive and negative affect	PANAS[f]	Pre/post each condition (acute); weeks 0, 5, 10, and 14 (chronic)	Secondary	10-50 per subscale; higher=greater affect	LMM
Frontal alpha and beta power	Emotiv EPOC X EEG system, 256 Hz	Pre/post each condition (acute); weeks 0, 5, 10, and 14 (chronic)	Secondary	Continuous; exploratory only	LMM
Mean HR[g], maximum HR, and %HRR[h]	Polar H10 chest strap monitor	3-min seated baseline before exercise; continuous during exercise	Secondary (motor load equivalence)	%HRR 0-100; equivalence margin prespecified	Paired comparisons and/or LMM
Rating of perceived exertion	Borg CR-10 scale	Immediately after exercise	Secondary (motor load equivalence)	0-10; higher=greater perceived exertion	Wilcoxon test and/or LMM
Accelerometry (CPM[i], vector magnitude, estimated steps)	Wrist-worn ActiGraph wGT3X-BT accelerometer	Continuous during exercise	Secondary (motor load equivalence)	Continuous; higher=greater external load	Paired comparisons and/or LMM
Presence	Igroup Presence Questionnaire	Postcondition (acute); the first and final intervention sessions (chronic phase)	Mediator	7-point scale; higher=greater presence; qualitative reference values	Exploratory mediation analysis
Cybersickness	Simulator Sickness Questionnaire	Immediately before and after VR[j] exposure (VR condition only)	Safety/control variable and potential moderator	Total severity; higher=greater symptoms; withdrawal threshold prespecified	Descriptive statistics and exploratory pre/post analyses
Enjoyment	Physical Activity Enjoyment Scale	Postcondition (acute); the first and final intervention sessions (chronic phase)	Mediator	16-80; higher=greater enjoyment	Exploratory mediation analysis
Participant expectations	Study-specific expectation item applied separately to each condition	Before the first experimental session only (baseline)	Moderator/control variable	0-10; higher=greater expected benefit	Exploratory moderation analysis

[a]CW: color-word score.

[b]CW′: expected color-word score.

[c]LMM: linear mixed model.

[d]EEG: electroencephalography.

[e]ICC: intraclass correlation coefficient.

[f]PANAS: Positive and Negative Affect Schedule.

[g]HR: heart rate.

[h]%HRR: percentage of heart rate reserve.

[i]CPM: counts per minute.

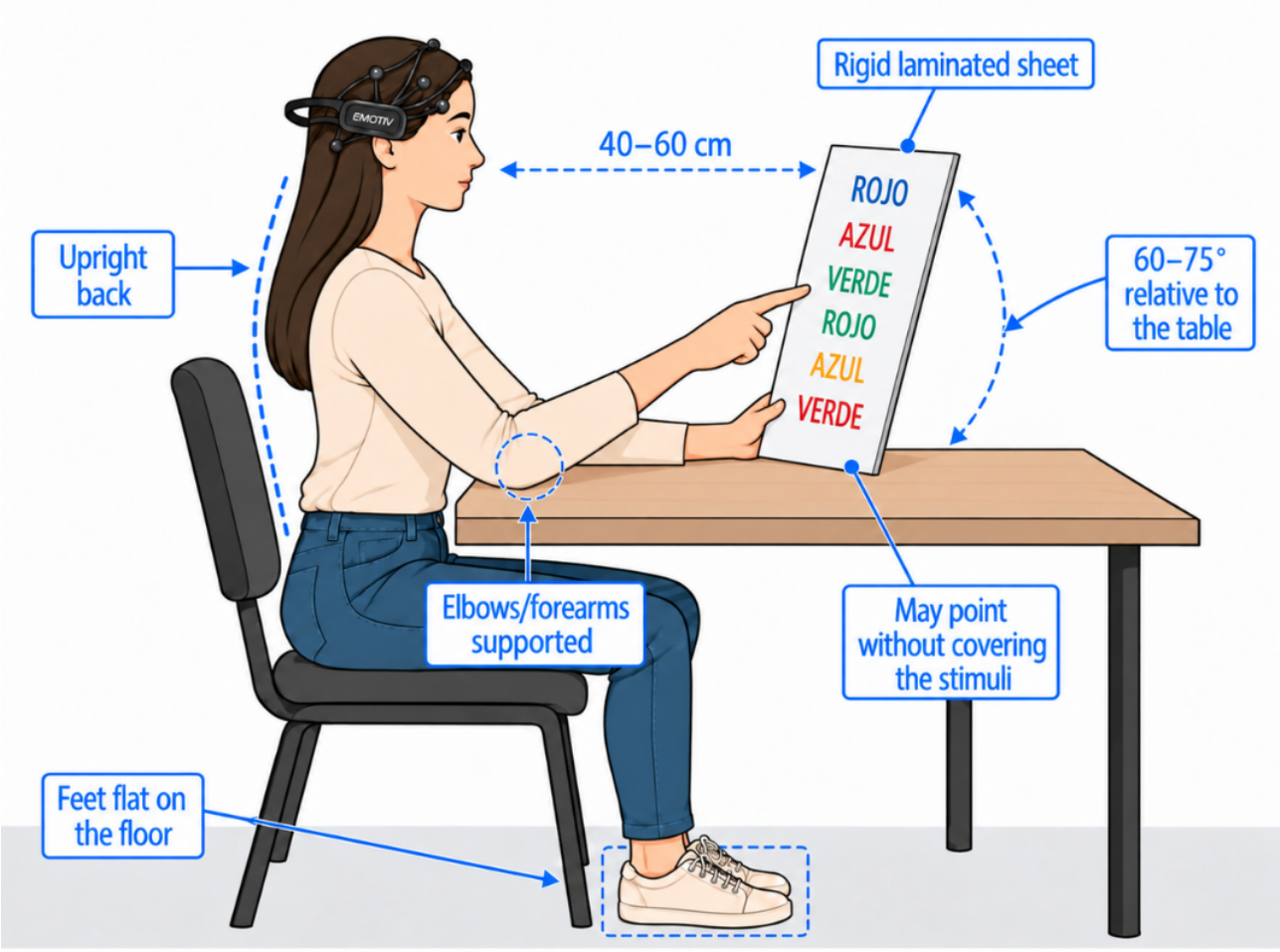
XSL•FO
RenderX

[j]VR: virtual reality.

Stroop Color and Word Test for Executive Function

Executive function will be assessed using the Stroop Color and Word Test in sheet-based format, composed of 3 conditions: word reading, color naming, and color-word interference [34,35]. In the interference condition, participants will be asked to name the color of the ink in which incongruent color words are printed [36-40]. The test will be administered in a quiet room following a standardized procedure. Participants will remain seated with an upright back, elbows and forearms supported, and feet flat on the floor. The laminated stimulus sheet will be positioned at an approximate viewing distance of 40 to 60 cm and at an angle of 60° to 75° relative to the table. Participants may point to the stimuli without covering them while wearing the Emotiv EPOC X electroencephalography system during recording, as shown in Figure 4. For each condition, the score obtained during the established 45-second administration time will be recorded. Based on the scores for words (W), colors (C), and color-word (CW), the expected color-word score will be calculated as CW′ = (W × C)/(W + C). The interference index will be calculated as CW – CW′ and will be considered the primary behavioral outcome of inhibitory control [35]. The index is a difference score and will be analyzed as a continuous variable; higher values indicate better resistance to interference.

Figure 4. Standardized participant position during administration of the Stroop Color and Word Test. The participant wears the Emotiv EPOC X electroencephalography system during recording. This schematic illustration was generated using a generative AI image tool based on specifications provided by the authors.



Electroencephalography Assessment of Frontal Cortical Activity

Recording Procedures

Electroencephalographic activity will be recorded using the wireless Emotiv EPOC X system (Emotiv Inc), a portable 14-channel electroencephalography device with saline-based wet sensors and 2 reference electrodes [41,42]. The system records at a sampling frequency of 256 Hz. Electrodes are positioned according to the international 10-20 system at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4, with CMS/DRL reference electrodes located in the parieto-occipital region. Electrode contact quality will be checked before each recording according to the manufacturer's recommendations, and sensors will be rehydrated with saline solution when necessary.

All electroencephalography recordings will be conducted under standardized conditions in a quiet room with controlled ambient illumination. To minimize potential confounding factors affecting cortical oscillatory activity, participants will be

XSL•FO
RenderX

instructed to avoid vigorous physical activity and alcohol consumption during the 24 hours prior to each assessment, and to avoid caffeine-containing beverages for at least 12 hours before testing. Whenever possible, electroencephalography assessments will be performed at the same time of day for each participant. Electroencephalography will be recorded only during assessment periods, and not during the exercise sessions, in order to reduce contamination from movement artifacts, sweating, headset vibration, and electromyographic activity. Each electroencephalography assessment will include a 3-minute eyes-open resting-state recording, during which participants will remain seated, relaxed, and focused on a fixed visual target. Electroencephalography will also be recorded during the Stroop Color and Word Test. Because the Stroop task will be administered in a sheet-based format, electroencephalography will be recorded continuously across the three 45-second conditions: word reading, color naming, and color-word interference. The start and end of each 45-second Stroop condition will be marked in real time using predefined keystroke event markers in EmotivPRO. The researcher will trigger the corresponding marker at the beginning and end of each condition, and these markers will be retained in the exported electroencephalography recording for subsequent condition segmentation and spectral analysis [43]. Electroencephalography data recorded during the Stroop task will be analyzed in the frequency domain, and spectral power in theta, alpha, and beta bands will be computed separately for each Stroop condition and averaged across artifact-free epochs.

Signal Preprocessing and Spectral Analysis

Electroencephalography data will be processed using EmotivPRO Analyzer software (Emotiv Inc). Signal quality will be monitored during data acquisition, and only recordings meeting the manufacturer's recommended quality criteria will be included in the analyses. The Emotiv EPOC X system was selected based on previous studies demonstrating acceptable validity and reliability for spectral electroencephalography analyses in cognitive and psychophysiological research settings [41,42]. Preprocessing will include artifact detection and rejection, robust rereferencing, high-pass filtering at 1 Hz, and low-pass filtering at 40 Hz. All recordings will also undergo visual inspection by an experienced investigator. Segments containing excessive movement artifacts, electrode instability, blink contamination, or substantial electromyographic interference will be excluded before spectral analysis. Channels exhibiting persistent poor signal quality throughout the recording will be excluded. An epoch will be defined a priori as contaminated if the signal exceeds ±100 µV at any included channel, or if it is identified as artifactual on visual inspection. Recordings containing more than 20% contaminated epochs after preprocessing, or fewer than 80% of channels meeting the contact-quality criterion, will be considered unreliable and excluded from the confirmatory analyses. These thresholds were established a priori by the research team rather than derived from a single published standard, as no consensus criteria exist for consumer-grade electroencephalography systems, and are reported in advance of data collection in accordance with current recommendations for transparent reporting of electroencephalographic methods [44].

Preprocessed electroencephalography signals will be transformed into the frequency domain using a fast Fourier transform. Spectral estimates will be computed using a sliding-window approach with a 256-sample Fourier window and a 64-sample step size, corresponding to 75% overlap. For resting-state analyses, continuous electroencephalography recordings will be segmented into consecutive 2-second epochs. Following artifact rejection, a minimum of 30 artifact-free epochs (≥60 seconds of valid electroencephalography data) will be required for inclusion in the final analyses. For Stroop-task recordings, electroencephalography data will be segmented into consecutive 2-second epochs within each of the 3 Stroop conditions: word reading, color naming, and color-word interference. Only artifact-free epochs will be retained for spectral analyses, and a minimum of 18 artifact-free epochs (≥36 seconds of valid electroencephalography data) will be required within a given condition for that condition to enter the analyses. Spectral power estimates in theta, alpha, and beta bands will be averaged across all valid epochs within each Stroop condition for each participant and assessment period. No event-related potential analyses will be conducted because the sheet-based administration does not allow precise stimulus-locked electroencephalography event marking.

Power spectral density will be calculated for conventional electroencephalography frequency bands, including theta (4-7 Hz), alpha (8-12 Hz), and beta (13-30 Hz) [26]. For each electrode and frequency band, both absolute power and relative power will be calculated.

Electroencephalography Outcome Measures

The confirmatory electroencephalographic end point is defined above as absolute frontal theta power (4-7 Hz) averaged across F3, F4, AF3, and AF4 during the color-word interference condition of the Stroop task. This end point was selected because frontal theta activity is closely related to cognitive control, attentional allocation, and inhibitory processing [33,45-48]. All other spectral outcomes are exploratory and are listed as such: relative frontal theta power; absolute and relative alpha (8-12 Hz) and beta (13-30 Hz) power over frontal electrodes [49,50]; all 3 bands during the word-reading and color-naming Stroop conditions; all 3 bands during the resting-state recording; and regional groupings outside the frontal region [51-53]. When appropriate, electrodes will be grouped into frontal, temporal, parietal, and occipital regions to improve interpretability and reduce the number of comparisons. Given the sheet-based administration of the Stroop task, event-related potentials such as N200 or P300 will not be analyzed, and electroencephalography activity during the task will be quantified using frequency-domain measures averaged across artifact-free epochs within each condition. To provide a direct estimate of measurement stability in this sample with this device, frontal theta power will additionally be recorded during the familiarization session, and the test-retest reliability of the confirmatory end point between the familiarization and preintervention recordings will be estimated as an intraclass correlation coefficient and reported. An observed change in the confirmatory end point will be interpreted as relevant to executive function only if it occurs in the prespecified band, region, and task condition; exceeds the prespecified smallest

XSL•FO
RenderX

effect of interest and the measurement error established by the reliability analysis; and is accompanied by change in the behavioral measure in the predicted direction. A statistically significant spectral change failing the second or third criterion will be reported as such and interpreted as inconclusive rather than as mechanistic support.

Hygiene Procedures

The Emotiv EPOC X system uses felt sensor pads, which will be replaced before use with each participant. The headset structure and electrode arms will be cleaned between participants using alcohol-based disinfectant wipes, following the manufacturer's recommendations and the hygiene and infection-control procedures established by the study center. Disposable materials will be discarded according to current institutional regulations.

Perceived Self-Efficacy

Perceived self-efficacy will be assessed using a 6-item scale developed according to Bandura's guidelines for constructing domain-specific self-efficacy scales, adapted to the context of active video games [54,55]. Items will be rated from 0 ("not at all capable") to 100 ("completely capable"). A global mean score will be calculated, with higher values indicating greater perceived self-efficacy. As a study-specific scale, scores will be analyzed as continuous variables, and internal consistency in this sample will be reported.

Positive and Negative Affect

Affective state will be assessed using the Spanish adaptation of the Positive and Negative Affect Schedule (PANAS) [56] based on the original scale developed by Watson et al [57]. The PANAS consists of 20 items grouped into two 10-item subscales: positive affect and negative affect. Participants will rate the extent to which they experience each affective state using a 5-point Likert scale, ranging from 1 ("not at all or very slightly") to 5 ("very much"). Separate scores will be calculated for positive and negative affect, with each subscale ranging from 10 to 50; higher scores indicate greater affect of the corresponding valence. Scores will be analyzed as continuous variables.

HR

HR will be continuously recorded using a Polar H10 chest strap monitor. Before each exercise session, participants will remain seated for 3 minutes to obtain a baseline HR recording. During the exercise session, mean HR, maximum HR, and percentage of HR reserve (%HRR) will be calculated using the Karvonen formula [58]: ([exercise HR – resting HR]/[estimated maximum HR – resting HR]) × 100. Estimated maximum HR will be calculated using the equation of Tanaka et al [31], 208 – 0.7 × age, since a maximal exercise test is not performed. %HRR will be the principal variable used to verify the equivalence of exercise intensity between conditions.

RPE

RPE will be assessed using the Borg CR-10 scale, with scores ranging from 0 to 10, where 0 represents "no exertion" and 10 represents "maximal exertion." The scale will be explained to participants before the intervention. RPE will be administered verbally at the end of each session, within the first minute after exercise completion. No cutoff will be applied; scores will be analyzed as a continuous indicator of perceived exertion and as a complement to the objective intensity measures described above.

Accelerometry

Triaxial accelerometry will be recorded using a wrist-worn ActiGraph wGT3X-BT accelerometer (ActiGraph LLC), configured to sample at 100 Hz and placed on the dominant wrist. Accelerometry will be used to characterize upper-limb movement and overall movement intensity during the intervention session, and to verify whether external motor load is comparable between conditions [28,59]. The accelerometer-derived variables will include activity counts per minute, vector magnitude, and mean acceleration. These variables were selected because they provide complementary information about movement volume, movement intensity, and overall external load during the task. Given the predominantly upper-body nature of Beat Saber, step count will be treated as an exploratory secondary variable rather than a primary indicator of task-related activity. Estimated energy expenditure during the session will additionally be derived from the accelerometry signal as a complementary objective indicator of external load. If relevant between-condition differences are observed, accelerometry-derived variables may be considered as covariates in the statistical analyses.

Habitual Physical Activity

Habitual physical activity level will be assessed at baseline using the International Physical Activity Questionnaire-Short Form (IPAQ-SF) [60]. The questionnaire will be used to estimate weekly physical activity expressed as metabolic equivalent of task (MET)-minutes/week, following the standard International Physical Activity Questionnaire (IPAQ) scoring protocol. In addition to the continuous MET-minutes/week value, participants will be classified into the low, moderate, and high activity categories defined by the official IPAQ scoring protocol. Both the continuous value and the category will be reported, and the continuous value will be used as the covariate in the statistical analyses.

Mediating and Moderating Variables of Immersion

To adequately interpret potential differences between VR and CS in cognitive and neurophysiological outcomes, variables related to the subjective experience of immersion will be assessed. These measures will help determine whether observed effects may be partly explained by perceived presence, enjoyment, or other psychological factors, rather than by the experimental condition alone.

Presence will be assessed using the Igroup Presence Questionnaire (IPQ), a 14-item instrument that measures the subjective experience of "being there" in a virtual or interactive environment, administered after each experimental condition [4]. IPQ items are rated on a 7-point scale and subscale means are computed. Scores will be analyzed as continuous variables; for descriptive purposes, scores will additionally be interpreted using the qualitative grading scale derived from a large reference database of IPQ responses [11].

XSL•FO
RenderX

Cybersickness will be assessed using the Simulator Sickness Questionnaire (SSQ), a 16-item instrument assessing symptoms of simulator sickness, including nausea, oculomotor discomfort, and disorientation [61]. The SSQ will be administered immediately before and after VR exposure, with the dual purpose of monitoring safety and quantifying its potential moderating effect on cognitive performance. For the SSQ, the total severity score is obtained by weighted summation of the nausea, oculomotor, and disorientation subscales, using the weights and conversion factor specified in the original scoring procedure [61]. No validated clinical cutoff exists for this instrument, and scores will be analyzed as continuous variables; higher scores indicate greater symptom severity. For descriptive purposes, scores will additionally be classified using the interpretive bands in common use in the VR literature, in which values below 5 indicate negligible symptoms, 5-10 minimal, 10-15 significant, 15-20 concerning, and above 20 problematic. These bands were developed to characterize simulator systems rather than to classify individual participants and are reported here as descriptive reference values only [62]. Because the questionnaire is administered both before and after exposure, both the raw postexposure score and the change from the pre-exposure baseline will be reported, and the interpretive bands will be applied to the raw postexposure score. In addition to stopping any session immediately if severe symptoms are reported, a postexposure total severity score above 20, or such a score on 2 consecutive sessions in the chronic study, will result in withdrawal of the participant from further immersive exposure. This threshold was established a priori by the research team as a safety criterion and is not presented as a validated clinical cutoff.

Enjoyment during physical activity will be assessed using the Spanish version of the Physical Activity Enjoyment Scale (PACES) [63]. This version includes 16 items rated on a 5-point Likert scale, ranging from 1 (“totally disagree”) to 5 (“totally agree”). A global enjoyment score will be calculated after reverse-scoring the negatively worded items. Total scores therefore range from 16 to 80, with higher scores indicating greater enjoyment. Scores will be analyzed as continuous variables. PACES will be administered after each experimental condition in the acute phase and after the first and final intervention sessions in the chronic phase.

Because previous beliefs about the expected benefits of VR or CS may influence participants’ responses, expectations will be assessed before the first experimental session using the same study-specific item applied separately to each condition and rated from 0 to 10. This will allow condition-specific expectations to be compared and considered as a potential moderating factor. The joint analysis of these variables will allow us to explore whether potential differences between conditions are mediated by subjective immersion and enjoyment or moderated by factors such as cybersickness and previous expectations.

The administration procedures for these instruments are detailed above. The main study variables and planned analyses are summarized in Table 1.

Procedure

Assessment and Monitoring Sequence

Acute Crossover Phase

Each participant will complete 2 experimental sessions, one in the VR condition and one in the CS, separated by a washout period of at least 72 hours. Each session will begin with a brief presession checklist to record relevant contextual information and verify participants’ readiness to exercise, including caffeine intake, recent medication use, previous physical activity, and dizziness or discomfort symptoms. At the first visit, written informed consent will also be obtained. Before exercising, participants will be fitted with the Polar H10 chest strap and will remain seated for 3 minutes to obtain a pre-exercise baseline HR recording. The self-efficacy scale and PANAS will then be administered, and only before the first experimental session, participant expectations will be recorded. Subsequently, the electroencephalography system will be placed, signal quality will be checked, and a 3-minute resting-state electroencephalography recording will be performed. The Stroop Color and Word Test will then be administered while participants wear the electroencephalography system. The wrist-worn accelerometer will be activated before the exercise task. The electroencephalography system will then be removed for the duration of the exercise task and refitted before the postexercise assessment, with contact quality reverified according to the manufacturer’s criteria prior to each recording.

The intervention task comprises 19 minutes 48 seconds of active gameplay and a total duration of approximately 23 minutes, including intersong transitions, and consists of the dance-based exergame task performed in the assigned condition, either VR or CS. HR and accelerometry will be recorded continuously during exercise, and RPE will be assessed verbally immediately after the session using the Borg CR-10 scale. Postexercise assessment will first include a 3-minute resting-state electroencephalography recording, followed by the Stroop Color and Word Test administered while participants continue wearing the electroencephalography system, in order to assess the immediate neurocognitive response after exercise. After electroencephalography recording, the self-efficacy scale and PANAS will be administered again. Presence (IPQ) and enjoyment (PACES) will be assessed after each experimental condition, while cybersickness (SSQ) will be assessed before and after VR exposure. Adverse events will be recorded at the end of each session.

Chronic Phase

The main outcome assessments will be conducted at baseline (week 0), at the midpoint of the intervention (week 5), postintervention (week 10), and 4-week follow-up (week 14). These assessments will include the Stroop Color and Word Test, electroencephalography, perceived self-efficacy, and PANAS, following the same standardized procedures described above. During the 10-week intervention, HR, accelerometry, RPE, adherence, and adverse events will be monitored at each training session. Enjoyment (PACES) and presence (IPQ) will be assessed after the first and final intervention sessions, and

XSL•FO
RenderX

cybersickness will be monitored before and after VR exposure when applicable.

Acute Phase: Randomized Crossover Design

In the acute phase, participants will be allocated in a 1:1 ratio to one of two condition sequences using a computer-generated random sequence produced with Research Randomizer: Sequence A, VR followed by CS, or Sequence B, CS followed by VR. Allocation will be stratified by sex so that sex is balanced across sequences. The allocation sequence will be generated by a researcher not involved in the delivery of the sessions or in outcome assessment and will be held in sequentially numbered sealed opaque envelopes opened only after eligibility has been confirmed and written informed consent obtained, so that allocation is concealed until assignment. Participants cannot be blinded to the assigned condition, for the reasons stated in the Limitations, but outcome data will be processed and analyzed by a member of the research team not involved in the delivery of the sessions or in outcome assessment. A minimum washout period of 72 hours will separate the 2 sessions, and the actual interval between sessions will be recorded for each participant and reported as a median and range rather than only as compliance with the minimum. Period and sequence effects will be estimated in the linear mixed models described below. We note explicitly that in a 2-sequence, 2-period design the differential carryover effect is aliased with the sequence effect and cannot be separated from it using within-period data alone; the sequence term is therefore not presented as a valid test of carryover, and 2 prespecified analyses are used instead, as described in the Statistical Analysis section.

Chronic Phase: RCT

In the chronic phase, participants will be randomized to either the VR or CS group using sex-stratified block randomization (blocks of 4). An independent researcher not involved in recruitment, intervention delivery, or outcome assessment will generate the allocation sequence and prepare sequentially numbered, sealed, opaque envelopes. The research team will enroll participants and reveal group allocation only after eligibility confirmation and written informed consent. The intervention will take place over 10 weeks, with 3 sessions per week, each comprising approximately 20 minutes of active gameplay and approximately 23 minutes in total, including intersong transitions, resulting in a total of 30 sessions per participant. At the beginning of each session, resting HR will be recorded for 3 minutes as a measure of pre-exercise physiological status. Formal assessments will be conducted at week 0 (preintervention), week 5 (midpoint), week 10 (postintervention), and week 14 (4-week follow-up), following the same standardized procedures described previously.

Adherence will be recorded at each session and calculated as the percentage of completed sessions out of the 30 planned sessions. Sessions that are started but not fully completed will be registered as partial sessions, and the reason for noncompletion or absence will be documented whenever possible. Participants completing at least 80% of the planned sessions will be considered adherent to the intervention. Adverse events, including cybersickness symptoms, dizziness, musculoskeletal discomfort, excessive fatigue, or any other unexpected event occurring during or immediately after the intervention, will be recorded at each session.

Following the intention-to-treat principle, participants who withdraw from the study will be included in the main analysis, with missing data handled through multiple imputation when the proportion of missing values exceeds 5% for a given variable [64].

Feasibility and Acceptability

Feasibility and acceptability are assessed against criteria established in advance. Recruitment feasibility is defined as achieving the target sample within the planned recruitment window. Retention is considered acceptable if 80% or more of randomized participants complete the final assessment. Intervention adherence is considered acceptable if participants complete 80% or more of the 30 planned sessions, as specified above. Data quality is considered acceptable if 80% or more of electroencephalographic recordings meet the prespecified signal-quality criteria. Tolerability is considered acceptable if withdrawal attributable to cybersickness occurs in fewer than 10% of participants exposed to the immersive condition. Acceptability is additionally assessed through the PACES, the SSQ, and the documented reasons for noncompletion of sessions. All criteria will be reported irrespective of whether they are met.

Statistical Analysis

Overview

All analyses will be performed using R (version ≥4.3.0) and the *lme4* package, with a significance level of α=.05 (2-tailed). The analysis will be conducted by a member of the research team who is not involved in the delivery of the intervention sessions or in the collection of outcome data. The prespecified statistical analysis plan, containing the model specifications, the estimands, the multiplicity structure, the handling of missing data and the sensitivity analyses, is provided as Multimedia Appendix 3. Analysis code will be made publicly available in an open repository at the time of results publication.

Primary Analysis

The primary confirmatory analysis concerns the Stroop interference index. For the acute crossover study, the model represents the full structure of the design, including the 2 periods and the 2 within-session measurements nested inside each period, and is specified in *lme4* notation as: outcome ~ condition * time + period + sequence + sex + habitual physical activity + (1 | participant) + (1 | participant: period). Here, condition is the exposure assigned in that period, time is the within-session assessment point (pre or post), period is the ordinal position of the session (first or second), and sequence is the randomized order to which the participant was allocated. Two random intercepts are specified: one for participant, capturing stable between-participant differences, and one for the participant-by-period combination, that is, the session, capturing the correlation between the pre and post measurements taken within the same session. Sex and habitual physical activity assessed with the IPAQ-SF are included as prespecified covariates, and the condition-by-time interaction is the term of

XSL•FO
RenderX

interest. For the chronic study, the corresponding model includes group, time as a factor with the planned assessment levels, the group-by-time interaction, the same covariates, and the baseline value of the outcome, with a random intercept for participant; the group-by-time interaction is the term of interest, and a prespecified contrast tests whether the between-group difference at week 10 exceeds that at week 5, which is the test of the cumulative-effect hypothesis. The week 14 follow-up will be analyzed exploratorily to examine whether postintervention differences are maintained after the end of the intervention.

Estimands

The estimands are specified following the framework of the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) E9(R1) addendum [65]. For the acute study: population, nonathletic adults aged 18 to 30 years meeting the eligibility criteria; variable, the Stroop interference index computed as CW – CW′; treatment condition, a single session of the assigned exposure; summary measure, the difference between conditions in the mean within-session change from preintervention to postintervention, adjusted for period, sequence, sex and habitual physical activity; intercurrent events, handled under a treatment policy strategy, with participants analyzed according to their randomized sequence irrespective of protocol deviations and a per-protocol analysis reported as a sensitivity analysis. In plain terms, the quantity to be estimated is the extent to which the immediate pre-to-post change produced by one condition exceeds that produced by the other in the same participants. For the chronic study, the estimand is the difference between groups in the change from baseline to week 10 in the same variable and population, under the same intercurrent-event strategy, with the week 5 assessment supporting the characterization of the intervention trajectory and the week 14 follow-up providing exploratory information on maintenance after intervention cessation.

Assessment of Period, Sequence, and Carryover Effects

Period and sequence are estimated as fixed effects in the model above. Because differential carryover is aliased with the sequence effect in a 2-sequence, 2-period design, 2 additional prespecified analyses are used to assess residual effects. First, the preintervention values of the second period are compared between sequences, adjusted for the preintervention values of the first period; if the effect of the first exposure had not dissipated before the second session, second-period baselines would differ systematically between sequences. This diagnostic is available only because both periods include a baseline measurement. Second, a sensitivity analysis restricted to first-period data is conducted and analyzed as a parallel-group comparison, which is unaffected by carryover by construction; if its estimate diverges materially from the crossover estimate, the first-period estimate will be given interpretative priority, and the divergence reported. The actual between-session interval will additionally be examined as a covariate in a sensitivity analysis.

Key Mechanistic and Secondary Analyses

The same modelling approach is applied to the confirmatory electroencephalographic end point and to the remaining secondary and exploratory outcomes. The brain-behavior relationship is prespecified rather than left to be decided after inspection of the data: the change in absolute frontal theta power during the color-word interference condition will be correlated with the change in the Stroop interference index, within condition and across the sample, with the correlation coefficient and its CI reported; and the exploratory mediation analysis is extended to test frontal theta change as a mediator of the effect of condition on the Stroop interference index, in addition to presence and enjoyment as previously specified. Prespecified secondary analyses will further (1) explore the influence of sex by including the time-by-condition-by-sex interaction; (2) examine order effects in the crossover study through the sequence-by-condition interaction, subject to the aliasing limitation stated above; and (3) characterize the trajectory of change in the chronic study across the planned assessment points. All mediation analyses are designated exploratory given the available sample sizes.

Multiple Comparisons Control

The confirmatory and exploratory analyses are separated a priori, and the family to which each correction applies is stated explicitly. The confirmatory family contains 2 end points tested hierarchically: the Stroop interference index, tested first at a 2-sided α of .05 without adjustment because it is a single prespecified end point for which the studies are powered; and the confirmatory electroencephalographic end point, tested second at the same α. No false discovery rate correction is applied within this family because of the hierarchical structure. The Benjamini-Hochberg false discovery rate correction, at a level of 5%, is applied separately within each of 3 prespecified exploratory families. Family A, secondary psychological outcomes: perceived self-efficacy, PANAS positive affect, and PANAS negative affect. Family B, exploratory electroencephalographic outcomes: relative frontal theta power; absolute and relative alpha and beta power over frontal electrodes; all 3 bands during the word-reading and color-naming Stroop conditions; all 3 bands during the resting-state recording; and regional groupings outside the frontal region. Family C, immersion-related mediators and moderators: presence, enjoyment, cybersickness, and expectations in the exploratory mediation and moderation analyses. Analyses of physiological and motor load equivalence are not included in any family because they are equivalence analyses rather than tests of superiority. Both unadjusted and adjusted *P* values will be reported for every analysis.

Equivalence of Physiological and Motor Load

Equivalence of exercise intensity between conditions will be evaluated using the two 1-sided tests procedure [66] rather than inferred from the absence of a significant difference, with percentage of HR reserve as the principal variable. The equivalence margin is set at 5 percentage points of HR reserve and is justified as the smallest difference in intensity that could plausibly produce a detectable difference in executive function on the basis of the acute exercise-cognition dose-response

XSL•FO
RenderX

literature. Mean HR, estimated energy expenditure, and accelerometry-derived load will be evaluated in the same way. If equivalence is not established for a given variable, that variable will be entered as a covariate in the primary model and a sensitivity analysis conducted with and without the adjustment.

Effect Size and Statistical Power

Cohen *d* (for comparisons between 2 conditions) and partial η_p^2 will be calculated for the linear mixed model effects. Interpretation: d<0.2 trivial; 0.2-0.5 small; 0.5-0.8 moderate; >0.8 large [67]. The observed statistical power will be reported for each primary analysis. A smallest effect of interest of d=0.5 is prespecified for the confirmatory electroencephalographic end point, corresponding to the effect size for which the studies are powered, so that a null result can be interpreted against a prespecified benchmark rather than as a mere absence of statistical significance. The intraclass correlation coefficient quantifying the test-retest reliability of that end point will be reported alongside it.

Missing Data

Missing data (dropout and session absences) will be handled through multiple imputation [64] when the proportion of missing values exceeds 5% for a given variable, following the intention-to-treat principle. The retention rate by group and session adherence rate will be reported.

Descriptive Statistics

Continuous variables will be presented as mean (SD) when normally distributed (verified using the Shapiro-Wilk test plus visual inspection of histograms and Q-Q plots) or as median and IQR when nonnormally distributed. Categorical variables will be expressed as absolute frequencies and percentages.

Ethical Considerations

The study will be conducted in accordance with Regulation (EU) 2016/679, the Spanish Organic Law 3/2018 on Personal Data Protection and Guarantee of Digital Rights, the Declaration of Helsinki, and the ethical principles applicable to research involving human participants. The protocol received a favorable opinion from the Research Ethics Committee of Universidad Loyola Andalucía on June 5, 2026 (approval code 260318 CE26666). Both studies are covered by a single prospective registration in the ISRCTN registry under ISRCTN15563829, registered on October 2, 2026. The registration record describes the acute crossover study and the chronic RCT separately, each with its own design, allocation procedure, target sample size, intervention schedule, outcome time points, and analysis plan, so that both are prospectively specified and individually identifiable within the record. No participant will be enrolled in either study before the registration record covering that study is publicly available; if registration or its amendment is delayed, the start of recruitment will be postponed accordingly so that prospective registration is preserved. All participants will receive detailed information about the aims, procedures, potential risks and benefits of the study, and will provide written informed consent voluntarily before participation. The participant information sheet and model informed consent form are provided as Multimedia Appendix 5. Relevant protocol amendments will be submitted to the Research Ethics Committee, updated in the trial registry when applicable, and reported in future publications.

Data will be coded using an alphanumeric identification system. The identification key will be stored in a password-protected file accessible only to the principal investigator. Digital data will be stored on institutional servers with restricted access, and physical documents will be kept in a locked cabinet. Data will be retained for a minimum of 5 years after study completion or publication of the results.

The risks associated with participation are considered minimal and may include mild muscle fatigue, transient discomfort related to the use of sensors (such as the chest strap or electroencephalography system), and possible dizziness or cybersickness associated with the use of VR headsets. To minimize these risks, all experimental sessions will be supervised by trained researchers with a background in physical activity and sport sciences and experience in exercise assessment, cognitive testing, and technology-supported interventions. The research team includes senior researchers with doctoral-level training, and all staff involved in data collection will receive standardized training in the study protocol before the first session. This training will cover delivery of the exercise task in both conditions according to the standardized script, fitting and use of the HR monitor and accelerometer, placement of the electroencephalography system and verification of signal quality according to the manufacturer's criteria, standardized administration and scoring of the Stroop Color and Word Test, procedures for safe use of the head-mounted display, and recognition and management of cybersickness or other adverse events. The study protocol was reviewed and endorsed by a consultant neurologist prior to submission to the research ethics committee, as required by the committee as a condition for approval. In addition, eligibility criteria will be checked before participation, and a basic first-aid kit, an automated external defibrillator, and a written emergency protocol will be available at the study site. Participants will be informed that they may withdraw from the study at any time without penalty. Cybersickness will be systematically monitored using the SSQ during VR sessions, and the session will be stopped immediately in the event of severe symptoms.

Results

This paper describes 2 studies that had not started recruitment at the time of submission. Ethics approval was obtained from the Research Ethics Committee of Universidad Loyola Andalucía on June 5, 2026 (approval code 260318 CE26666). Both studies were prospectively registered in the ISRCTN registry under a single record (ISRCTN15563829, registered on October 2, 2026), which describes the acute crossover study and the chronic RCT separately.

Recruitment for both studies is expected to begin in October 2026. Data collection is anticipated to be completed by December 2027, and the first results are expected by June 2028. No participant had been enrolled, and no data had been collected at the time of submission; therefore, no outcome data are available, and no analyses have been performed.

XSL•FO
RenderX

Discussion

Hypothesized Findings

Based on the available evidence, we hypothesize that the immersive condition will produce greater improvement in the primary outcome, the Stroop interference index, than the matched video-guided comparator. We further hypothesize greater absolute frontal theta power in the confirmatory end point, which would support the interpretation of a neurophysiological substrate underlying changes in inhibitory control. Higher perceived self-efficacy, enjoyment, and positive affect are also anticipated in the immersive condition. Physiological and motor load, assessed through HR, percentage of HR reserve, RPE, and accelerometry, are expected to be equivalent between conditions within the prespecified margin. These statements are hypotheses derived from prior literature and from the rationale of the present design; they are not results, and they will be confirmed or refuted only after data collection is complete. Should the hypothesized differences be observed, they will be interpreted as effects of the immersive exergame as an integrated package of immersion, interactivity, and real-time performance feedback, and not as isolated effects of immersion.

Principal Findings

This protocol addresses several limitations identified in previous research on exergames and immersive VR, including limited experimental control, insufficient standardization of motor tasks, and the lack of objective neurophysiological measures and immersion-related variables. The inclusion of presence, cybersickness, enjoyment, participant expectations, and positive and negative affect will allow the study to explore potential mechanisms underlying differences between conditions, helping to determine whether the observed effects are related to the combined immersive and interactive characteristics of VR, to motor or physiological load, or to subjective psychological and affective responses. In addition, the systematic monitoring of motor load through accelerometry, HR, and RPE will strengthen the internal validity of the design. The reformulation of the comparison as the immersive exergame considered as an integrated package vs a matched video-guided comparator is deliberate: it states what the design manipulates rather than what would be desirable to manipulate, and it restricts the claims made to those the design can support. Dissociating immersion from interactivity and performance feedback requires a design with an additional interactive nonimmersive condition and is identified here as the necessary next study rather than as something the present protocol achieves.

Comparison With Prior Work

Previous meta-analytic evidence indicates that exergames may exert moderate effects on executive function, particularly on inhibition and cognitive flexibility, although the available evidence is methodologically heterogeneous [3]. Studies of immersive VR have associated this technology with greater activation of frontal networks and enhanced motivation [7-10], but few have integrated neurophysiological and behavioral measures under controlled conditions [17-19] and even fewer have incorporated the variables required to interpret the mechanisms underlying the effects observed [8,68]. The 2 randomized trials most directly comparable to the present protocol are those of Huang [20], which compared immersive with nonimmersive exergaming over 4 weeks in adults older than 50 years, and Lachowicz et al [21], which compared immersive VR training with a no-training control condition in amateur esports athletes. Neither used an active comparator matched for musical and motor content, neither examined acute and chronic responses within a single protocol, neither combined behavioral and neurophysiological outcomes with a prespecified brain-behavior link, and neither measured presence, cybersickness, enjoyment, or expectations, which means that a positive finding in either cannot be attributed to a specific mechanism. By matching songs, tempo, movement sequence, and difficulty across conditions, by verifying the equivalence of physiological and motor load, and by recording portable electroencephalography during standardized assessments with a single prespecified confirmatory end point, the present protocol is designed to address these gaps.

Limitations

Some limitations should be acknowledged. First, electroencephalography will be recorded using a portable system rather than a laboratory-grade high-density device, which may reduce spatial resolution and increase susceptibility to noise [16,41]. Electroencephalography will be recorded only during resting state and the Stroop task, not during exercise, as movement, sweating, headset vibration, and electromyographic activity could introduce substantial artifacts; this limits the ability to capture dynamic cortical activity during the exergame.

Second, participants will be recruited from a single educational center using convenience sampling, which may limit generalizability to the broader population of young adults. Third, participants cannot be blinded to the experimental condition, which may introduce expectation-related bias; this will be partially addressed by assessing expectations before the first session and considering them as a potential moderator [3].

Fourth, and most importantly, although the 2 conditions are matched in song order, active gameplay duration, tempo, difficulty, and expected motor pattern, they differ concurrently in immersion, interactivity, real-time performance feedback, and goal structure. The design therefore does not permit any observed difference to be attributed to immersion, to interactivity, or to performance feedback separately: these features are manipulated as a package, and the results will be reported and interpreted on that basis. An interactive 2D comparator would improve equivalence of interactivity but, in the absence of a 2D version of the task used, would require a different game and would degrade the equivalence of musical and motor content; the comparator was chosen to preserve the latter, and the corresponding limitation is accepted explicitly rather than minimized [8].

Fifth, participants are cognitively healthy young adults at or near the peak of executive functioning, so an intervention effect may be difficult to detect even if present. Three features of the design mitigate this: the Stroop interference index is a difference score sensitive to variation within the normal range rather than a screening instrument with a ceiling; participants with regular

XSL•FO
RenderX

experience of rhythm-based VR exergames are excluded; and the acute study uses a within-participant design, which removes between-participant variance and increases sensitivity to small changes. Sixth, in the 2-sequence, 2-period crossover, differential carryover is aliased with the sequence effect and cannot be separated from it; the prespecified second-period baseline comparison and the first-period-only sensitivity analysis mitigate but do not eliminate this constraint. Finally, the temporal scope of the chronic study is bounded: it estimates the effect of a 10-week program and characterizes the trajectory across the planned assessments. Although a 4-week follow-up is included, the temporal scope of the chronic study remains limited, and longer-term maintenance beyond week 14, adherence, and transfer to habitual physical activity are not evaluated. These limitations are common in controlled exercise and VR intervention studies and were anticipated in the study design.

Conclusions

By combining strict experimental control, objective electroencephalography measures and the systematic assessment of potential mediators and moderators of VR-related effects, together with psychological and affective outcomes, this protocol is expected to provide relevant evidence for the development of technology-supported exercise interventions aimed at cognitive health and physical education settings.

Acknowledgments

The authors acknowledge the Industrial Doctorate Program of the University of Seville and the Spanish Institute for Global Education for their support. The authors also thank the staff and students of Santa Joaquina de Vedruna School (Seville, Spain) for their collaboration in the development of this study protocol. A complete declaration of generative AI use structured according to the GAIDeT (Generative AI Delegation Taxonomy) framework is provided in Multimedia Appendix 4. All authors reviewed, revised, and approved every part of the manuscript and accept full responsibility for its content.

Funding

SO-G is supported by grant JDC2024-053581-I, funded by MICIU/AEI/10.13039/501100011033 and by the European Social Fund Plus (ESF+). The funders had no role in the study design, conduct, analysis, reporting, or decision to publish.

Data Availability

Data sharing does not apply to this article, as no datasets were generated and analyzed during the preparation of this study protocol. The datasets generated and analyzed during the subsequent study will be available from the corresponding author on reasonable request.

Generative AI Use

Generative AI tools were used in the preparation of this manuscript, as follows. ChatGPT (OpenAI) was used for language editing and to improve the clarity of the English text. Claude (Anthropic) was used to create the flow diagrams and the study schedule figure (Figures 1-3) based on content, structure, and instructions fully specified by the authors. ChatGPT (OpenAI) was also used to generate Figure 4 from an author-provided description of the standardized participant position defined in the protocol. No generative AI tool was used to generate the scientific content of the manuscript, to define the research question or the methodological design, to select the outcomes or instruments, to develop the statistical analysis plan, or to interpret findings; no data have been collected at this stage, and no tool was used in any analysis.

Authors' Contributions

Conceptualization: MB-A, SO-G, AF-M
Methodology: MB-A, SO-G, BS, AF-M
Project administration: MB-A, AF-M
Supervision: BS, AF-M
Writing–original draft: MB-A, SO-G
Writing–review & editing: MB-A, SO-G, BS, AF-M
All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

SPIRIT 2025 checklist.
[PDF File (Adobe PDF File), 275 KB-Multimedia Appendix 1]

XSL•FO
RenderX

Multimedia Appendix 2

TIDieR checklist.
[PDF File (Adobe PDF File), 994 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Statistical analysis plan.
[PDF File (Adobe PDF File), 336 KB-Multimedia Appendix 3]

Multimedia Appendix 4

GAIDeT declaration.
[PDF File (Adobe PDF File), 260 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Informed consent form.
[PDF File (Adobe PDF File), 106 KB-Multimedia Appendix 5]

References

1. Diamond A. Executive functions. Annu Rev Psychol. 2013;64:135-168. [FREE Full text] [doi: 10.1146/annurev-psych-113011-143750] [Medline: 23020641]
2. Hillman CH, Erickson KI, Kramer AF. Be smart, exercise your heart: exercise effects on brain and cognition. Nat Rev Neurosci. 2008;9(1):58-65. [doi: 10.1038/nrn2298] [Medline: 18094706]
3. Stanmore E, Stubbs B, Vancampfort D, de Bruin ED, Firth J. The effect of active video games on cognitive functioning in clinical and non-clinical populations: a meta-analysis of randomized controlled trials. Neurosci Biobehav Rev. 2017;78:34-43. [FREE Full text] [doi: 10.1016/j.neubiorev.2017.04.011] [Medline: 28442405]
4. Schubert T, Friedmann F, Regenbrecht H. The experience of presence: factor analytic insights. Presence (Camb). 2001;10(3):266-281. [doi: 10.1162/105474601300343603]
5. Slater M. Immersion and the illusion of presence in virtual reality. Br J Psychol. 2018;109(3):431-433. [doi: 10.1111/bjop.12305] [Medline: 29781508]
6. Mocco A, Valmaggia L, Bernardi L, Alfieri M, Tarricone I. Enhancing physical activity with immersive virtual reality: a systematic review. Cyberpsychol Behav Soc Netw. 2024;27(5):303-317. [doi: 10.1089/cyber.2023.0394] [Medline: 38506442]
7. Doulou F, Piolino P, Angeard N. Virtual reality programs targeting executive functions and social cognition evaluation and/or rehabilitation in children with ADHD or ASD-A narrative review. Front Psychol. 2025;16:1583052. [FREE Full text] [doi: 10.3389/fpsyg.2025.1583052] [Medline: 41267765]
8. Makransky G, Petersen GB. The Cognitive Affective Model of Immersive Learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. Educ Psychol Rev. 2021;33(3):937-958. [doi: 10.1007/s10648-020-09586-2]
9. Bird JM, Karageorghis CI, Baker SJ, Brookes DA, Nowicky AV. Ready exerciser one: effects of music and virtual reality on cycle ergometer exercise. Br J Health Psychol. 2021;26(1):15-32. [doi: 10.1111/bjhp.12445] [Medline: 32538512]
10. Park JH. Detection of mild cognitive impairment using a virtual reality-based Stroop task: a cross-sectional study of embodied behavioral markers. J Neuroeng Rehabil. 2025;22(1):176. [FREE Full text] [doi: 10.1186/s12984-025-01714-9] [Medline: 40783531]
11. Melo M, Gonçalves G, Vasconcelos-Raposo J, Bessa M. How much presence is enough? Qualitative scales for interpreting the Igroup Presence Questionnaire score. IEEE Access. 2023;11:24675-24685. [doi: 10.1109/access.2023.3254892]
12. Wang D, Peng Y, Haddouk L, Vayatis N, Vidal P. Assessing virtual reality presence through physiological measures: a comprehensive review. Front Virtual Real. 2025;6:1530770. [doi: 10.3389/frvir.2025.1530770]
13. Ojeda de Ocampo F, Hernández-Melgarejo G, Ramírez-Treviño A, Fuentes-Aguilar RQ. Presence assessment in virtual reality: a systematic literature review. Appl Sci (Basel). 2026;16(6):3102. [doi: 10.3390/app16063102]
14. Tan E, Troller-Renfree SV, Morales S, Buzzell GA, McSweeney M, Antúnez M, et al. Theta activity and cognitive functioning: integrating evidence from resting-state and task-related developmental electroencephalography (EEG) research. Dev Cogn Neurosci. 2024;67:101404. [FREE Full text] [doi: 10.1016/j.dcn.2024.101404] [Medline: 38852382]
15. Castanho L, Martinho DV, Saial AC, Gouveia BR, Gouveia ÉR, Ribeiro F. The efficacy of virtual reality-based EEG neurofeedback in health-related symptoms relief: a systematic review. Appl Psychophysiol Biofeedback. 2025;50(4):613-633. [doi: 10.1007/s10484-025-09730-0] [Medline: 40715707]
16. Ayuso-Moreno R, Rubio-Morales A, Durán-Rufaco A, García-Calvo T, González-Ponce I. EEG-based assessment of mental fatigue in students: a systematic review of measurement methods and data processing protocols. Appl Sci (Basel). 2025;16(1):234. [doi: 10.3390/app16010234]

XSL•FO
RenderX

17. Han JC, Zhang C, Cai YD, Li YT, Shang YX, Chen ZH, et al. Neuroimaging features for cognitive fatigue and its recovery with VR intervention: an EEG microstates analysis. Brain Res Bull. 2025;221:111223. [FREE Full text] [doi: 10.1016/j.brainresbull.2025.111223] [Medline: 39864596]
18. Park H, Hwang H, Kim M, Park H, Jeon I, Shon YM, et al. Electroencephalogram changes after virtual reality-based cognitive behavioral therapy for panic disorder. J Korean Med Sci. 2025;40(31):e185. [FREE Full text] [doi: 10.3346/jkms.2025.40.e185] [Medline: 40795342]
19. Wu J, Guan Z, Li C, Wang W, Sun R, Zhu Z, et al. Brain functional connectivity after Stroop task induced cognitive fatigue. Sci Rep. 2025;15(1):22342. [FREE Full text] [doi: 10.1038/s41598-025-08330-6] [Medline: 40596555]
20. Huang KT. Exergaming executive functions: an immersive virtual reality-based cognitive training for adults aged 50 and older. Cyberpsychol Behav Soc Netw. 2020;23(3):143-149. [doi: 10.1089/cyber.2019.0269] [Medline: 31794673]
21. Lachowicz M, Klichowski M, Serweta-Pawlik A, Rosciszewska A, Żurek G. Effects of short- and long-term virtual reality training on concentration performance and executive functions in amateur esports athletes. Virtual Real. 2025;29(2):91. [doi: 10.1007/s10055-025-01161-w]
22. Chan AW, Boutron I, Hopewell S, Moher D, Schulz KF, Collins GS, et al. SPIRIT 2025 statement: updated guideline for protocols of randomised trials. BMJ. 2025;389:e081477. [doi: 10.1136/bmj-2024-081477] [Medline: 40294953]
23. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: Template for Intervention Description and Replication (TIDieR) checklist and guide. BMJ. 2014;348:g1687. [FREE Full text] [doi: 10.1136/bmj.g1687] [Medline: 24609605]
24. Chang YK, Labban JD, Gapin JI, Etnier JL. The effects of acute exercise on cognitive performance: a meta-analysis. Brain Res. 2012;1453:87-101. [doi: 10.1016/j.brainres.2012.02.068] [Medline: 22480735]
25. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332. [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
26. Cohen MX. Analyzing Neural Time Series Data: Theory and Practice. Cambridge, Massachusetts. MIT Press; 2014.
27. Ludyga S, Gerber M, Brand S, Holsboer-Trachsler E, Pühse U. Acute effects of moderate aerobic exercise on specific aspects of executive function in different age and fitness groups: a meta-analysis. Psychophysiology. 2016;53(11):1611-1626. [doi: 10.1111/psyp.12736] [Medline: 27556572]
28. Impellizzeri FM, Marcora SM, Coutts AJ. Internal and external training load: 15 years on. Int J Sports Physiol Perform. 2019;14(2):270-273. [doi: 10.1123/ijspp.2018-0935] [Medline: 30614348]
29. Halson SL. Monitoring training load to understand fatigue in athletes. Sports Med. 2014;44 Suppl 2(Suppl 2):S139-S147. [FREE Full text] [doi: 10.1007/s40279-014-0253-z] [Medline: 25200666]
30. Achten J, Jeukendrup AE. Heart rate monitoring: applications and limitations. Sports Med. 2003;33(7):517-538. [doi: 10.2165/00007256-200333070-00004] [Medline: 12762827]
31. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol. 2001;37(1):153-156. [FREE Full text] [doi: 10.1016/s0735-1097(00)01054-8] [Medline: 11153730]
32. Borg G. Borg's Perceived Exertion and Pain Scales. Champaign, IL. Human Kinetics; 1998.
33. 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]
34. Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol. 1935;18(6):643-662. [doi: 10.1037/h0054651]
35. Golden CJ. Stroop Color and Word Test. Chicago, IL. Stoelting; 1978.
36. Appelbaum LG, Boehler CN, Davis LA, Won RJ, Woldorff MG. The dynamics of proactive and reactive cognitive control processes in the human brain. J Cogn Neurosci. 2014;26(5):1021-1038. [FREE Full text] [doi: 10.1162/jocn_a_00542] [Medline: 24345171]
37. Bianco V, Berchicci M, Mussini E, Perri RL, Quinzi F, Di Russo F. Electrophysiological evidence of anticipatory cognitive control in the Stroop task. Brain Sci. 2021;11(6):783. [FREE Full text] [doi: 10.3390/brainsci11060783] [Medline: 34199201]
38. Koga S, Rothermel R, Juhász C, Nagasawa T, Sood S, Asano E. Electrocorticographic correlates of cognitive control in a Stroop task-intracranial recording in epileptic patients. Hum Brain Mapp. 2011;32(10):1580-1591. [FREE Full text] [doi: 10.1002/hbm.21129] [Medline: 20845393]
39. Xiao Y, Chou CC, Cosgrove GR, Crone NE, Stone S, Madsen JR, et al. Cross-task specificity and within-task invariance of cognitive control processes. Cell Rep. 2023;42(1):111919. [FREE Full text] [doi: 10.1016/j.celrep.2022.111919] [Medline: 36640346]
40. Tafuro A, Ambrosini E, Puccioni O, Vallesi A. Brain oscillations in cognitive control: a cross-sectional study with a spatial Stroop task. Neuropsychologia. 2019;133:107190. [FREE Full text] [doi: 10.1016/j.neuropsychologia.2019.107190] [Medline: 31513806]
41. Badcock N, Mousikou P, Mahajan Y, de Lissa P, Thie J, McArthur G. Validation of the Emotiv EPOC(®) EEG gaming system for measuring research quality auditory ERPs. PeerJ. 2013;1:e38. [FREE Full text] [doi: 10.7717/peerj.38] [Medline: 23638374]
42. Williams NS, McArthur GM, de Wit B, Ibrahim G, Badcock N. A validation of Emotiv EPOC Flex saline for EEG and ERP research. PeerJ. 2020;8:e9713. [FREE Full text] [doi: 10.7717/peerj.9713] [Medline: 32864218]

XSL•FO
RenderX

43. Keystroke markers. EmotivPRO V30 Doc. 2026. URL: https://emotiv.gitbook.io/emotivpro-v3/event-markers/keystroke-markers [accessed 2026-10-27]
44. Pernet C, Garrido MI, Gramfort A, Maurits N, Michel CM, Pang E, et al. Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. Nat Neurosci. 2020;23(12):1473-1483. [doi: 10.1038/s41593-020-00709-0] [Medline: 32958924]
45. Hanslmayr S, Pastötter B, Bäuml K-H, Gruber S, Wimber M, Klimesch W. The electrophysiological dynamics of interference during the Stroop task. J Cogn Neurosci. 2008;20(2):215-225. [FREE Full text] [doi: 10.1162/jocn.2008.20020] [Medline: 18275330]
46. Rostami F, Esteki A, Sarabi-Jamab A. Interference control under the Stroop task and brain oscillatory activity among internet addicts compared to non-addicts. Basic Clin Neurosci. 2024;15(5):631-648. [doi: 10.32598/bcn.2023.4974.1] [Medline: 40583887]
47. Zhao Y, Tang D, Hu L, Zhang L, Hitchman G, Wang L, et al. Concurrent working memory task decreases the Stroop interference effect as indexed by the decreased theta oscillations. Neuroscience. 2014;262:92-106. [doi: 10.1016/j.neuroscience.2013.12.052] [Medline: 24406438]
48. Chinn LK, Pauker CS, Golob EJ. Cognitive control and midline theta adjust across multiple timescales. Neuropsychologia. 2018;111:216-228. [FREE Full text] [doi: 10.1016/j.neuropsychologia.2018.01.031] [Medline: 29410123]
49. Arakaki X, Hung S, Rochart R, Fonteh AN, Harrington MG. Alpha desynchronization during Stroop test unmasks cognitively healthy individuals with abnormal CSF amyloid/Tau. Neurobiol Aging. 2022;112:87-101. [FREE Full text] [doi: 10.1016/j.neurobiolaging.2021.11.009] [Medline: 35066324]
50. Zhu Y, Wu D, Sun K, Chen X, Wang Y, He Y, et al. Alpha and theta oscillations are causally linked to interference inhibition: evidence from high-definition transcranial alternating current stimulation. Brain Sci. 2023;13(7):1026. [FREE Full text] [doi: 10.3390/brainsci13071026] [Medline: 37508958]
51. Proskovec AL, Wiesman AI, Wilson TW. The strength of alpha and gamma oscillations predicts behavioral switch costs. Neuroimage. 2019;188:274-281. [FREE Full text] [doi: 10.1016/j.neuroimage.2018.12.016] [Medline: 30543844]
52. Min BK, Park HJ. Task-related modulation of anterior theta and posterior alpha EEG reflects top-down preparation. BMC Neurosci. 2010;11:79. [FREE Full text] [doi: 10.1186/1471-2202-11-79] [Medline: 20584297]
53. Herrmann CS, Demiralp T. Human EEG gamma oscillations in neuropsychiatric disorders. Clin Neurophysiol. 2005;116(12):2719-2733. [doi: 10.1016/j.clinph.2005.07.007] [Medline: 16253555]
54. Bandura A. Self-Efficacy: The Exercise of Control. New York. W. H. Freeman; 1997.
55. Bandura A. Guide for constructing self-efficacy scales. In: Pajares F, Urdan T, editors. Self-Efficacy Beliefs of Adolescents. Greenwich, CT. Information Age Publishing; 2006:307-337.
56. López-Gómez I, Hervás G, Vázquez C. Adaptación de las "escalas de afecto positivo y negativo" (PANAS) en una muestra general española. Behav Psychol Psicol Conduct. 2015;23(3):529-548. [FREE Full text]
57. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol. 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
58. Karvonen MJ, Kentala E, Mustala O. The effects of training on heart rate: a longitudinal study. Ann Med Exp Biol Fenn. 1957;35(3):307-315. [Medline: 13470504]
59. Boots B, Berg D, Hewitt E, Naugle K, Naugle K. Physical activity and enjoyment in active virtual reality games in youth: comparative analysis of Gorilla Tag and Beat Saber. JMIR Serious Games. 2025;13:e66593. [FREE Full text] [doi: 10.2196/66593] [Medline: 40168642]
60. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
61. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol. 1993;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
62. Bimberg P, Weissker T, Kulik A. On the usage of the simulator sickness questionnaire for virtual reality research. 2020. Presented at: IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); 2020 March 22-26:464-467; Atlanta, GA, USA. [doi: 10.1109/vrw50115.2020.00098]
63. Moreno JA, González-Cutre D, Martínez C, Alonso N, López M. Propiedades psicométricas de la Physical Activity Enjoyment Scale (PACES) en el contexto español. Estud Psicol (Madr). 2008;29(2):173-180. [doi: 10.1174/021093908784485093]
64. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York. Wiley; 1987.
65. ICH E9(R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. 2019. URL: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf [accessed 2026-10-27]
66. Lakens D. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. Soc Psychol Personal Sci. 2017;8(4):355-362. [FREE Full text] [doi: 10.1177/1948550617697177] [Medline: 28736600]
67. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, N.J. Erlbaum Associates; 1988.

XSL•FO
RenderX

68. Kourtesis P, Collina S, Doumas LAA, MacPherson SE. Technological competence is a pre-condition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. Front Hum Neurosci. 2019;13:342. [FREE Full text] [doi: 10.3389/fnhum.2019.00342] [Medline: 31632256]

Abbreviations

%HRR: percentage of heart rate reserve
BPM: beats per minute
CONSORT: Consolidated Standards of Reporting Trials
CS: comparator condition
GAIDeT: Generative AI Delegation Taxonomy
HR: heart rate
ICH: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IPAQ: International Physical Activity Questionnaire
IPAQ-SF: International Physical Activity Questionnaire-Short Form
IPQ: Igroup Presence Questionnaire
ISRCTN: International Standard Randomised Controlled Trial Number
MET: metabolic equivalent of task
PACES: Physical Activity Enjoyment Scale
PANAS: Positive and Negative Affect Schedule
RCT: randomized controlled trial
RPE: rating of perceived exertion
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
SSQ: Simulator Sickness Questionnaire
TIDieR: Template for Intervention Description and Replication
VR: virtual reality

Edited by J Sarvestan; submitted 30.Jun.2026; peer-reviewed by G Ochi, M Shimaoka; comments to author 17.Aug.2026; revised version received 06.Oct.2026; accepted 23.Oct.2026; published 12.Nov.2026

Please cite as:
Becerra-Alonso M, Ortega-Gómez S, Sañudo B, Feria-Madueño A
Immersive Virtual Reality Exergaming vs Video-Guided Screen-Based Exercise for Cognitive and Neurophysiological Adaptations in Young Adults: Protocol for Acute and Chronic Experimental Studies
JMIR Res Protoc 2026;15:e105801
URL: https://www.researchprotocols.org/2026/1/e105801
doi: 10.2196/105801
PMID:

©Miguel Becerra-Alonso, Sonia Ortega-Gómez, Borja Sañudo, Adrián Feria-Madueño. Originally published in JMIR Research Protocols (https://www.researchprotocols.org), 12.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Research Protocols, is properly cited. The complete bibliographic information, a link to the original publication on https://www.researchprotocols.org, as well as this copyright and license information must be included.

XSL•FO
RenderX