City Walk: Embodied Locomotion Improves Route Efficiency and Spatial Memory in a Virtual City

Paolo Boffi[1,2*], Davide Tonsi[1], Nermin Mina[2], Alberto Gallace[2],
Pier Luca Lanzi[1]

[1]Department of Electronics, Informatics, and Bioengineering (DEIB),
Polytechnic University of Milan, Milan, Italy.
[2]Mind and Behavior Technological Center (MiBTec), University of
Milan-Bicocca, Milan, Italy.

*Corresponding author(s). E-mail(s): paolo.boffi@polimi.it;
Contributing authors: davide.tonsi@mail.polimi.it;
n.mina@campus.unimib.it; alberto.gallace1@unimib.it;
pierluca.lanzi@polimi.it;

Abstract

Spatial navigation is a relevant target for assessment and training that relies on the interplay between egocentric route execution and allocentric, map-like representations. Virtual Reality (VR) enables ecologically plausible navigation tasks under experimental control, yet outcomes can strongly depend on immersion and locomotion interfaces that determine the availability of self-motion cues.

In this paper, we present *City Walk*, a VR serious game designed to support implicit training and assessment of spatial navigation in a urban environment. The experience begins with guided acclimatization and gradually shifts to unguided exploration, time pressure, obstacle-induced re-planning, and ends with a map-based landmark placement task. *City Walk* implements two interaction modalities: Desktop VR (DVR) and Enhanced-Immersive VR (E-IVR), which combines a Head-Mounted Display (HMD) with an omnidirectional treadmill.

We report a pilot between-subjects study comparing the two conditions. The protocol comprises five navigation levels with increasing demands, as well as a landmark placement test on an overhead map, supported by in-app logging and gaze-based landmark observation in the E-IVR build. E-IVR yielded substantially higher route efficiency and improved map-based landmark placement, while level completion times tended to be longer. User-centered questionnaires indicated comparable usability and tolerability across conditions.

Keywords: Immersive Virtual Reality, Spatial Navigation, Embodied Locomotion, Eye Tracking

1 Introduction

Spatial navigation is a fundamental cognitive function that allows people to orient in the environment, plan routes, and reach goals in both familiar and unfamiliar settings (Chan et al. 2012; Serino et al. 2025). Navigation relies on the interplay between egocentric representations (e.g., stimulus–response route execution) and allocentric, map-like representations that encode relationships among landmarks and locations (Klatzky 1998; Ekstrom et al. 2014; Iglói et al. 2009). Importantly, difficulties in navigation and landmark memory are frequently observed in aging and in several neurological conditions, including dementia and focal brain injuries, where they can translate into reduced autonomy and quality of life (Chan et al. 2012; Cogné et al. 2017). This makes the assessment and training of spatial navigation – and, in particular, the promotion of flexible allocentric strategies – a relevant target for preventive and rehabilitative interventions.

Virtual reality (VR) offers a compelling methodological framework for this goal because it enables realistic large-scale environments under tight experimental control, with high repeatability, safe deployment, and rich behavioral logging (Kelly and Gibson 2007; Diersch et al. 2019). VR has been used both for intervention-oriented training (with evidence of transfer to real-world navigation) and for mechanistic studies aimed at isolating how different cues and strategies support learning (Taube et al. 2013; Clemenson et al. 2020; Hejtmanek et al. 2020). In parallel, serious games provide an additional layer of structure and motivation, supporting sustained engagement and scalable training protocols while preserving the possibility of fine-grained data collection (Mortara et al. 2014; Perdomo and Pérez 2025). In navigation-relevant domains, serious-game approaches have shown promising effects on spatial memory and allocentric remapping in urban-like virtual settings, including in populations with cognitive vulnerability (Sacco et al. 2022; Abd-Alrazaq et al. 2022).

A key open issue, however, is that *how* users move through a Virtual Environment (VE) is not a neutral implementation detail. Locomotion interfaces determine which self-motion cues (visual, vestibular, proprioceptive) are available and can therefore affect spatial learning, workload, and cybersickness (Slater 2009; Diersch et al. 2019; Lackner 2014; Clifton and Palmisano 2020). Many applications rely on controller-based translation and rotation, which can increase movement throughput but may reduce congruent bodily cues; conversely, more embodied solutions such as omnidirectional treadmills can provide richer sensorimotor contingencies but introduce practical constraints and potential cognitive–motor interference (Warren and Bowman 2017; Hooks et al. 2020; Wang et al. 2023). These interaction choices are particularly relevant when the intended use case involves repeated sessions and older adults, for whom usability, fatigue, balance demands, and tolerance may constrain adoption (Diersch et al. 2019).

In this work, we introduce *City Walk*, a VR serious game designed to support implicit training and assessment of spatial navigation in a compact, everyday-like urban environment. The game implements a fixed multi-level progression that gradually shifts from guided acclimatization to unguided exploration, time pressure, obstacle-induced re-planning, and transfer from a novel start location, culminating in a map-based landmark placement task. In the present study, the same city and level logic are implemented in two builds that differ in immersion and locomotion modality: a desktop baseline (DVR) and an enhanced immersive build (E-IVR) that combines an HMD with an omnidirectional treadmill to increase embodied sensorimotor involvement. To support process-level analyses, *City Walk* includes modular in-app logging of movement heatmaps and route-efficiency metrics, performance in the map task, and (in E-IVR) gaze-based landmark observation, enabling links between online behavior and offline spatial-memory outcomes.

The present manuscript reports a pilot evaluation comparing DVR and E-IVR in young adults, with two primary aims: to validate the feasibility of the protocol and logging pipeline, and to obtain preliminary evidence on whether embodied immersive locomotion yields measurable advantages for navigation efficiency and survey-level spatial knowledge relative to a desktop baseline. In addition, we collect user-centered measures (workload, user experience, usability, and cybersickness) to assess acceptability and inform iterative refinement prior to larger studies in the intended middle-aged/older adult population.

2 Background and Related Work

2.1 Spatial Navigation and Cognitive Maps

Spatial navigation can be defined as the ability to move purposefully through the environment by selecting and following routes that connect different locations in space (Chan et al. 2012; Smith et al. 2024; Park et al. 2024). It is a fundamental cognitive function, that enables individuals to orient themselves and reach goals in both familiar and unfamiliar environments (Serino et al. 2025). Difficulties in spatial navigation are frequently observed in aging and in several neurological conditions, including dementia and focal brain injuries, where they may lead to loss of autonomy and reduced quality of life (Chan et al. 2012; Cogné et al. 2017).

Humans and other animals rely on multiple strategies to represent space and guide navigation, but two main frames of reference are typically distinguished: *allocentric* and *egocentric* representations (Ekstrom et al. 2014). Allocentric representations encode the spatial relationships between objects independently of the observer's current position, supporting the formation of cognitive maps that can be used to plan novel routes and to reorient after displacements (Klatzky 1998). This strategy is strongly linked to the hippocampus, which plays a central role in encoding and retrieving allocentric spatial information (Iglói et al. 2009). In contrast, egocentric representations encode the position of objects relative to the observer's body, such as "on the left" or "in front of me". Egocentric navigation can rely on stimulus–response rules (e.g., "turn right at the corner") or on memorized sequences of body turns; simple egocentric strategies are primarily supported by the striatum, whereas more complex sequential strategies can also involve the hippocampus (Klatzky 1998; Iglói et al. 2009).

A central component of spatial navigation is the use of landmarks – salient cues in the environment such as buildings, natural features or other distinctive objects (Chan et al. 2012; Yesiltepe et al. 2021; Jabbari et al. 2022). Visual landmarks are especially important and can be classified along several dimensions. A common distinction is between *proximal* and *distal* landmarks: proximal landmarks are located close to the observer and typically support egocentric strategies, but they can be less reliable because they can be transient, movable, or occluded; distal landmarks, such as tall buildings or mountains, are visible from many locations and provide stable reference information that facilitates the construction of allocentric maps (Chan et al. 2012). More specifically, landmarks can be described as *beacons*, which directly mark the goal location, *orientation cues*, supporting heading and directional decisions, *associative cues*, linked to specific actions along a route, and *reference frames*, boundaries or contours that define the overall spatial layout, such as coastlines or city walls (Chan et al. 2012).

2.2 Virtual Reality and Serious Games for Spatial Navigation

Virtual reality (VR) has been increasingly adopted as a tool to investigate and train spatial navigation because it allows researchers to reproduce realistic yet highly controlled environments (Jonson et al. 2021; Clemenson et al. 2020). Virtual Environments (VEs) can be flexibly modified to test specific factors while providing safe and accessible setups even for individuals with reduced mobility or cognitive impairment (Capriotti et al. 2025). VR also enables the collection of rich, multimodal data, including brain activity, behavioral responses, navigation strategies and longitudinal changes in performance (Diersch et al. 2019; Kelly and Gibson 2007). In this context, VR has been proposed as a technology that may *"redefine experimental practice in neuroscience and psychology"*, particularly for spatial navigation research (Diersch et al. 2019, p. 2).

Work in VR navigation broadly spans two complementary lines: intervention-oriented studies that use VR to train navigation skills and evaluate transfer to real-world performance, and mechanistic or assessment-oriented studies that use VEs to examine navigation behavior and its neural underpinnings under controlled conditions (Taube et al. 2013; Diersch et al. 2019; Hejtmanek et al. 2020; Clemenson et al. 2020).

Early evidence for training effectiveness was provided by Bliss et al. (1997), who reported that firefighters trained in a virtual building (via a Head-Mounted Display (HMD) and mouse) performed a subsequent real-world search-and-rescue task faster and more accurately than a no-training control group. More recently, Xu et al. (2025) compared active VR exploration, passive video training, and no training in older adults using a digital replica of a real building. Participants in the active VR condition later navigated the corresponding real environment with reduced time and distance – even on new routes – supporting transfer from virtual to real-world settings. In rehabilitation, Kober et al. (2013) implemented route-finding training using a non-immersive VR system for neurologic patients with spatial disorientation and reported improvements in route learning and performance on neuropsychological measures after repeated sessions. Other training work has targeted cognitive map formation specifically: McLaren-Gradinaru et al. (2020) used repeated navigation sessions in a non-immersive 3D video game set in a fictional city (*Centerville*) and reported preliminary evidence of improved allocentric spatial memory on standard tasks.

A substantial body of research uses VR primarily as an experimental platform to study spatial cognition and navigation strategies. Virtual implementations of classical paradigms (e.g., the Morris Water Maze, Radial Arm Maze, and geometric arenas) facilitate large-scale testing and enable translational comparisons with animal research at relatively low cost (Kelly and Gibson 2007; Thornberry et al. 2021). For example, Iglói et al. (2009) used a virtual star maze to show that sequential egocentric and allocentric strategies can be acquired in parallel, by manipulating starting positions and available cues to dissociate which representation participants had learned. A systematic review by Cogné et al. (2017) further showed that VR-based navigation tasks are sensitive to aging, Alzheimer-type dementia, brain injury and schizophrenia, and can reveal specific deficits in landmark memory, route learning and environmental configuration that are not always captured by traditional neuropsychological tests.

Within this broader context, serious games provide a complementary framework for delivering spatial navigation training and assessment. Serious games are video games designed primarily to educate or train players rather than to provide entertainment alone (Mortara et al. 2014; Theodoropoulos and Antoniou 2022; Michael and Chen 2006). They have been widely used to enhance cognitive functions such as attention, memory and spatial navigation, with promising outcomes across domains (Oei and Patterson 2013; Ackerman et al. 2010; Perdomo and Pérez 2025). A bibliometric analysis reported that serious games have grown across fields since 2019, with cognition-related studies occupying a particularly prominent position in relevance and development (Perdomo and Pérez 2025). Serious games also frequently integrate virtual or augmented reality technologies, supporting interactivity, motivation, high repeatability of experimental protocols, and continuous collection of in-game performance data (Checa and Bustillo 2020; Perdomo and Pérez 2025; Lampropoulos and Kinshuk 2024).

Empirical evidence supports the potential of serious games for navigation-related outcomes. Sacco et al. (2022) developed *MindTheCity!*, a 3D game set in a fictional town composed of multiple districts, where players retrieve bicycle components distributed across the environment. Across three experiments with repeated sessions, they observed improvements in several aspects of spatial memory, consistent with enhanced remapping of space in allocentric coordinates. More broadly, a systematic review and meta-analysis by Abd-Alrazaq et al. (2022) indicated that serious games can be effective and safe tools for improving cognitive abilities in older adults with cognitive impairment, supporting their use in populations likely to benefit from engaging, technology-mediated interventions. These findings complement earlier evidence that non-action video games can produce specific cognitive improvements and that older adults can learn and use such games effectively – an important prerequisite for aging-focused research (Oei and Patterson 2013; Ackerman et al. 2010).

Taken together, the literature suggests that VR and serious games form a strong methodological combination for spatial navigation research and cognitive training. VR provides controlled yet realistic environments and enables fine-grained behavioral and physiological data collection, while serious games offer engaging task structures and scalable training protocols (Diersch et al. 2019; Perdomo and Pérez 2025). This combination appears particularly promising for designing implicit or explicit procedures aimed at improving allocentric navigation strategies and landmark-based spatial memory (Diersch et al. 2019; Sacco et al. 2022; Ellis 2009).

2.3 Locomotion and Immersion in VR Navigation

Immersion in VR is classically defined by the extent to which a system can *replace* the user's sensory access to the physical world with computer-generated stimulation, thereby hiding (or substantially occluding) the surrounding reality from the user's view (Riva and Waterworth 2014; Slater and Sanchez-Vives 2016). Under this perspective, VR systems are commonly grouped into three categories based on their level of perceptual isolation and user envelopment: *non-immersive* systems (e.g., desktop displays; Desktop Virtual Reality, DVR), which preserve full awareness of the real environment; *semi-immersive* systems (e.g., large-screen or projection-based setups), which partially surround the user and enhance engagement without fully occluding reality; and *immersive* systems (typically HMD- or CAVE-based; Immersive Virtual Reality, IVR), which provide stereoscopic, head-tracked views that largely hide the physical surroundings and support a first-person experience of the VE (Jensen and Konradsen 2018; Cipresso et al. 2018; Di Natale et al. 2020). Beyond this display-centric taxonomy, and coherently with previous literature (Radianti et al. 2020; Boffi et al. 2024), we use the term *Enhanced Virtual Reality* (EVR) to denote VR setups that extend sensorimotor involvement beyond vision (and basic audio) by integrating additional interfaces that provide more realistic interaction and/or feedback (e.g., haptic/force-feedback devices or locomotion platforms). When EVR is implemented with an IVR (HMD-based) setup in our study, we refer to it as *Enhanced-IVR* (E-IVR).

However, in navigation contexts immersion depends not only on display technology but also on how users move through the VE (Slater 2009; Hořejší et al. 2025). Many VR navigation applications rely on HMDs combined with hand-held controllers or simple input devices such as joysticks or keyboards (Monteiro et al. 2025; Steed and Lai 2025; Buttussi and Chittaro 2021). In these setups, users typically remain seated or standing in place while translating and rotating through the environment via button presses or joystick movements. The headset provides stereoscopic vision and, when equipped with six degrees of freedom (6 DoF), tracks both head rotations and translations; integrated 3D audio can further contribute to a sense of presence (Lee et al. 2017). However, because the body does not physically move, self-motion is conveyed primarily through visual and auditory cues. The resulting mismatch between visual motion and the absence of congruent vestibular and somatosensory input is a key mechanism underlying cybersickness (Lackner 2014; Clifton and Palmisano 2020). The literature therefore emphasizes the importance of maintaining high and stable frame rates – often on the order of 75 frames per second or higher – to preserve motion smoothness and reduce sensory conflict (Lee et al. 2017; Chang et al. 2020).

Driving simulators represent one approach to reintroducing vestibular information into VR locomotion. Beyond steering wheel and pedal interaction, advanced simulators can integrate motion platforms that reproduce weight shifts during acceleration,

Fig. 1: An example of Omnidirectional treadmill: the Virtualizer, by Cyberith.

braking, and turns, providing motion cues that target the vestibular system and meaningfully influence steering and speed control (Gabes and Mühlberger 2024). Visual information can be delivered via single or multi-monitor setups or via HMDs; recent work suggests that, for some driving tasks, there is no clear evidence of strong differences in immersion between multi-monitor and HMD-based configurations, and professional simulators therefore often do not require HMDs (Gabes and Mühlberger 2024). Nonetheless, even with motion platforms, these systems cannot fully replicate the range of accelerations experienced in real driving, and residual inconsistencies between visual and bodily cues can still induce motion sickness (Lackner 2014). Despite these limitations, driving simulators remain highly valuable for research questions that depend on vestibular stimulation (e.g., speed perception, steering control).

Omnidirectional treadmills (ODTs) extend this principle by enabling users to physically walk in any direction while remaining in place (Figure 1). Typical ODTs include a low-friction walking surface together with a ring-and-harness system that stabilizes balance and keeps the user centered; motion is then captured through sensor systems that estimate walking direction and speed (and sometimes body rotation) to drive locomotion in the VE (Diersch et al. 2019; Warren and Bowman 2017). These systems are used in gaming, professional training, simulations, and rehabilitation, and they can improve performance and precision in navigation tasks by coupling visual information with more natural locomotor patterns (Diersch et al. 2019; Chakraborty et al. 2024). However, ODTs are constrained by cost and space requirements and may introduce a dual-task situation in which motor demands compete with cognitive processing (Hooks et al. 2020; Wang et al. 2023). Evidence from VR navigation studies with older adults suggests that treadmill-based locomotion can sometimes impair performance in this population because walking and cognitive processing draw on shared resources (Diersch et al. 2019).

Overall, locomotion interfaces for VR navigation span a continuum from purely controller-based movement to solutions that engage vestibular and somatosensory systems (e.g., driving simulators and ODTs). The available literature suggests that higher immersion and richer bodily feedback can support learning, presence and ecological validity (Gutiérrez et al. 2007; Fusco and Tieri 2022), but may increase complexity, provoke cybersickness, and – in certain target groups such as older adults – heighten cognitive-motor interference (Diersch et al. 2019; Lackner 2014). Locomotion choice is therefore a central design variable in VR navigation studies because it determines the cue set available to the navigator and the practical trade-offs among immersion, usability, and safety.

2.4 Eye Tracking and Landmark Use in Virtual Navigation

Eye tracking offers a direct behavioral window onto how navigators sample visual information while moving through an environment. Modern VR headsets can integrate infrared cameras that track pupil and corneal reflections in real time (Moreno-Arjonilla et al. 2024). This approach – known as video oculography (van der Geest and Frens 2002) – enables estimation of gaze origin and direction concurrently with head pose and position. Earlier eye-tracking systems relied on intrusive magnetic contact lenses (van der Geest and Frens 2002; Robinson 1963), whereas current infrared-based solutions are non-invasive and can be embedded into HMDs, enabling the study of attention, perception and decision-making during more naturalistic navigation tasks (Moreno-Arjonilla et al. 2024; Callahan-Flintoft et al. 2021).

From a spatial cognition perspective, eye tracking is particularly informative because visual landmarks are primary cues used to build allocentric and egocentric representations of space (Klatzky 1998; Chan et al. 2012). As outlined in subsection 2.3, landmarks may be proximal or distal, and Chan et al. (2012) proposed a functional taxonomy (beacon, orientation, associative, and reference-frame cues). Because different landmark types support different navigational functions, patterns of gaze allocation can provide indirect evidence about the navigator's strategy – for example,whether behavior is dominated by local beacons or instead emphasizes distal orientation and reference-frame cues (Chan et al. 2012).

VR research also underscores the central role of landmarks when other cues are reduced, which motivates measuring how attention is allocated to them during navigation. Youngstrom and Strowbridge (2012), for example, used a virtual water maze in which mice (head-fixed on a levitated spherical treadmill) had access only to visual information from a surrounding display; animals learned to navigate to a hidden platform based solely on the configuration of visual landmarks, suggesting that allothetic visual cues can be sufficient to support spatial learning in VR. Related VR paradigms further indicate that nonvisual information can complement visual processing when visual cues are impoverished: Gröhn et al. (2005) reported that participants could succeed in a VR gate-navigation task using auditory cues, although performance was best with combined audiovisual signals. Together, such findings emphasize that landmark processing is often central when navigation relies heavily on vision and navigation performance may reflect how visual and nonvisual information is sampled and integrated – making eye tracking a valuable tool when visual information is expected to be task-relevant.

IVR can increase ecological validity by allowing participants to actively explore large-scale environments while experimenters record position, head orientation, and – when available – eye movements at high temporal resolution (Diersch et al. 2019; Zhao et al. 2023). Eye tracking in these settings supports gaze-based metrics such as total dwell time on landmarks and the number of sustained fixations above a given temporal threshold. These measures can be related to subsequent performance on explicit landmark recognition, route recall, or map reconstruction tasks (Viaene et al. 2016; Zhao et al. 2023; Wenczel et al. 2017). In principle, longer and more frequent fixations on distal orientation or reference-frame cues may be consistent with allocentric map construction, whereas heavier reliance on nearby beacons may align with egocentric or sequential strategies (Klatzky 1998; Chan et al. 2012). In clinical contexts, this approach is further motivated by evidence that VR navigation tasks are sensitive to landmark and route-memory deficits in aging and dementia (Cogné et al. 2017), highlighting the potential value of quantifying how visual information is sampled during navigation.

3 Materials and Methods

3.1 *City Walk*

3.1.1 Application Design

City Walk is a serious game designed to support implicit training of spatial navigation in an everyday-like urban environment. The application targets a broad user base, from young to elderly adults, with a particular focus on individuals who are not habitual video game players. For this reason, the interaction model was deliberately kept simple: the core task consists exclusively of moving from a starting location to a destination within a small virtual city, without additional mini-games or complex interactions. The game is intended to be easy to learn and to reuse in different experimental protocols in the domain of cognitive training.

The design goal was to manipulate immersion and locomotion while keeping the cognitive demands of the navigation task constant. To this end, the same city layout, level structure, and game logic are shared across two builds that differ only in interaction modality and degree of sensorimotor involvement: a DVR version, used as control condition, and an E-IVR version, combining an HMD with an omnidirectional treadmill (the Cyberith Virtualizer[1]). This dual-build architecture enables direct comparison between traditional keyboard–mouse navigation and embodied locomotion on a motion platform within the same virtual environment.

Both builds are implemented in Unity (version 2021.3.24f1), sharing the same underlying scene structure, level definitions, and event-driven game logic. In the DVR version, the player navigates the city using a standard first-person controller. View orientation is controlled with the mouse, while translation is mapped to keyboard arrows

---

[1] https://www.cyberith.com/

Fig. 2: City environment, seen from the height.

or WASD keys. Completion and failure of each level are communicated via dedicated end-of-level screens, and transitions between levels are mediated by a loading screen that conveys contextual information about the upcoming task. This configuration provides a baseline navigation condition with low sensorimotor immersion but familiar interaction and minimal hardware requirements. In the E-IVR version, the same navigation tasks and feedback screens are presented through an HTC Vive Pro Eye HMD while the player avatar is controlled via an omnidirectional treadmill. Controller-based translation is disabled, and forward motion and turns are driven by the user's physical walking and body rotations on the Virtualizer. Visual and auditory feedback remain identical to the DVR version, ensuring that performance differences can be attributed primarily to the locomotion modality and immersion-related sensorimotor contingencies rather than to changes in task content.

Across both versions, the application maintains the same narrative framing, starting/ending point logic, and level sequence (described in the following subsections). This design ensures that differences in performance and behavior can be attributed primarily to the locomotion modality and immersion level rather than to changes in game content or task demands.

3.1.2 City Layout and Landmark Design

The VE consists of a small coastal city located on an island, designed to balance ecological plausibility and experimental control (Figure 2). Both versions of the application use the same city scene, so that navigation demands and spatial information are identical across conditions. The same city layout is available in both daytime and nighttime scenes, allowing us to probe spatial learning under different lighting conditions while preserving an identical map structure. The creation of the city map followed three main steps: definition of the street topology, placement of salient landmarks, and filling the remaining space with generic buildings to control visibility. All design choices

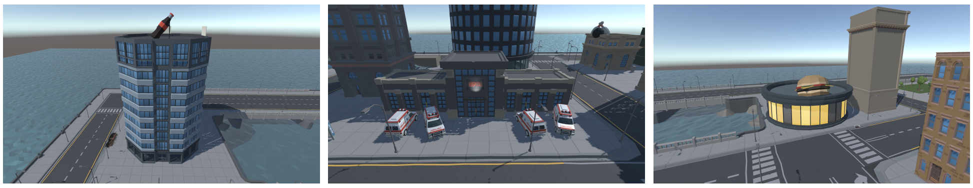


Fig. 3: Three different landmarks. Left: the office (*Core*); middle: the hospital (*Intermediate*); right: the fast food (*Optional*).

were oriented to support the player's learning process while maintaining a moderate level of complexity.

At the topological level, the city island has an approximately rectangular shape when viewed from above, with deliberate irregularities in each quadrant. The initial design consisted of a regular grid with four vertical and four horizontal streets, a format often used in navigation experiments (Smith et al. 2024; Starrett et al. 2021; König et al. 2019). This grid was then modified by introducing curved segments and local asymmetries, especially near the coastline, so that different areas of the city are easier to distinguish and remember. In particular, the southeastern–central sector preserves a more regular grid pattern, whereas the upper and left sections contain more curved and irregular streets. The southwestern quadrant includes a recessed piece of terrain that differentiates it from the other corners and is used in the game as the neighborhood of the player's home.

To ensure compatibility with the Cyberith Virtualizer used in the E-IVR condition, the entire city is modeled on a flat terrain: no uphill or downhill roads are present, and most streets are two-lane. Despite the absence of vertical relief, the morphology of the road network provides additional orientation cues. Three distinctive bridges are placed on the north-western side: a pair of "twin bridges" and one longer bridge, together with a three-lane road running along the northern edge of the city and a single-lane road in the southwest corner. These features act as secondary references that can be recognized visually and, in the case of the bridges, also through audio cues (via a looping seagull sound), supporting acclimatization without resorting to explicit maps. The city is entirely surrounded by water, which serves as a global reference frame for large-scale orientation.

Buildings are divided into generic constructions and distinctive landmarks. Generic buildings are used to fill the terrain and block long-distance lines of sight, reproducing the occlusion conditions of a real city. Without these occluders, participants would have an unrealistically easy view of all salient buildings and could orient themselves using only a few highly distinctive cues. Landmark buildings, by contrast, are key components for orientation: their placement and appearance are central to the intended learning process. Landmarks were first positioned according to the street topology and only afterwards were the remaining spaces filled with generic houses, apartment blocks and skyscrapers. Some of these generic structures are specifically placed to hide important landmarks (e.g., the player's house and office) from many viewpoints, forcing participants to rely on indirect cues to locate them. Distinctive landmarks were selected among buildings that are easily interpretable by a wide range of users (e.g., city hall, office, police station, hospital, park, fast food restaurant, sushi restaurant, bowling alley). Most of them are associated with a characteristic audio clip that is played at regular intervals in their vicinity, such as seagulls at the bridges, a telephone ringing at the office, a bowling ball hitting pins at the bowling alley, grill sounds at the fast food, and the chirping of birds in the park. In the night scene, all landmarks are also illuminated to create a highlight effect when walking around the city. Depending on their function and recognizability, landmarks are grouped into complexity tiers (Figure 3). *Core* landmarks are used as start or end points in the early levels (e.g., city hall, office, home) and act as beacons: they are explicitly mentioned in the instructions and are expected to be encoded rapidly. *Intermediate* landmarks (e.g., police station, hospital, bowling alley, park) serve as waypoints along routes and associative cues during the guided acclimatization phase. *Optional* landmarks (e.g., fast food, sushi) are never mentioned explicitly and can blend into the neighborhood, making their discovery and use more demanding. The spatial arrangement of the landmarks is designed to favor the construction of a cognitive map rather than simple stimulus–response chains. Core landmarks are located at or near the corners of the city (e.g., city hall in the northeast, office in the northwest, home in the southwest), so that their mutual relations can be inferred using the coastline as a reference. Some landmarks are visible from many points in the city (e.g., the office as a tall distal landmark) and act as orientation cues, whereas others are primarily local beacons whose audio and visual features are only available when the player is nearby (e.g., bowling alley, fast food, sushi). Figure 4 shows the placement of the various landmarks in the city environment. This variety of proximal and distal cues allows participants to adopt different strategies and supports allocentric learning: the same landmark can function as proximal or distal depending on the current position, and the network of routes connecting them admits multiple paths between common start–goal pairs. Finally, the distribution of streets and landmarks is coordinated with the level design (described in the next subsection) to discourage reliance on a single fixed route. For example, the home is located in a recess in the southwest area and is visually masked by skyscrapers; it must typically be reached by combining knowledge about other landmarks rather than by direct sight. Obstacles are later introduced along the most direct connection between home and office, blocking the long bridge that constitutes the shortest path. This forces players to exploit alternative routes and to use landmark relationships and street morphology to reorient themselves, encouraging the adoption of more flexible and allocentric navigation strategies over the course of the game.

3.1.3 Task Structure and Levels

City Walk is organized as a fixed sequence of six levels: five in-situ navigation tasks within the virtual city, followed by a map-based spatial memory task. The same level structure is used across versions. For each level, a configuration asset specifies the scene to be loaded, the identifiers of the start and end locations, whether the level is timed and the initial countdown value, and whether obstacles are enabled, ensuring consistent behavior across builds and experimental runs.

The first level introduces the participant to the city and serves as an acclimatization phase. The character has just moved to the town and starts at a bus stop in front of

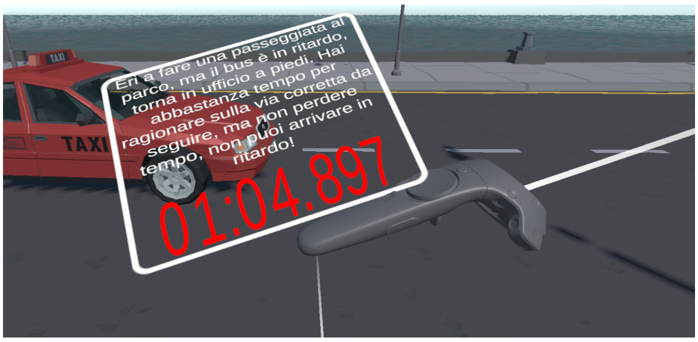
Fig. 4: Placement of all the notable landmarks in the city environment. *Core* landmarks are highlighted in red, *Intermediate* landmarks in yellow, *Optional* landmarks in green.

the city hall during daytime, with the objective of reaching home. Along this route, the player passes by several salient landmarks (police station, park, hospital, bowling alley) before arriving at the house, which is used as the endpoint. To reduce initial task difficulty and to standardize exposure to key landmarks, this level provides stepwise guidance via textual hints. Hints are triggered at predefined locations along the route and are displayed as text at the top of the screen in the DVR condition and on the VR watch in the E-IVR condition.

In the second level, the participant starts from the office at the end of the workday and must find the way home without any GPS-like guidance. The scene now takes place at night, and no countdown is active. The design goal is to encourage exploration of the city and active use of visible landmarks to plan a route. The absence of time pressure gives participants sufficient time to reason about alternative paths, compare options at intersections, and consolidate an internal representation of the spatial relations between the office, home, and intermediate landmarks.

The third level starts from the participant's home in the morning: the character is late for work and must reach the office within a limited time. The same daytime city scene is used, but this level is configured as a timed task with a countdown of 90 seconds, displayed on the in-game chronometer (Figure 5). This level is designed to induce moderate time pressure while remaining solvable within the allowed time, thereby encouraging participants to reuse the route learned in Level 2 but in the opposite direction. The task thus specifically tests the ability to exploit previously acquired landmarks and routes under temporal constraints.

In the fourth level, the participant again starts from the office at night and must return home, as in Level 2, but parts of the city are blocked due to construction work. Obstacles are enabled in the level configuration so that specific road segments



Eri a fare una passeggiata al parco, ma il bus è in ritardo, torna in ufficio a piedi. Hai abbastanza tempo per ragionare sulla via corretta da seguire, ma non perdere tempo, non puoi arrivare in ritardo!

Fig. 5: VR chronometer, with the time remaining displayed in red.

become impassable, forcing the player to abandon any purely egocentric “memorized turn sequence” and to identify an alternative route. The level is untimed, in order to focus on route re-planning and on the use of known landmarks (e.g., the city hall) as intermediate targets. The design explicitly aims to assess whether participants rely on allocentric knowledge of the city layout when previously used paths are no longer available.

The fifth level places the participant at the bus stop near the park on the eastern coast during a lunch break. The bus to return to the office has broken down, and the player must walk to the office within two minutes. This level combines a novel starting position (never used as start or goal in previous navigation levels) with a time limit that is long enough to allow strategic reasoning but still generates pressure. The design requires participants to integrate knowledge gained in earlier levels to localize themselves relative to known landmarks and to select an efficient route to the office. The route naturally exposes players to specific intermediate landmarks (e.g., the fast-food restaurant) that later appear in the map-based task, supporting subsequent assessment of landmark learning.

The sixth and final level is a map-based scene set inside the player’s house. The participant stands in a room with a large overhead map of the city on the wall and a set of post-it notes on a table, each representing one of the landmarks encountered during navigation (Figure 6). The map contains more placeholders than notes (12 placeholders for 8 landmarks), which prevents solving the task by simple exclusion strategies. Participants must place each note on the map at the location they believe corresponds to the landmark’s position. Notes can only be placed on predefined slots, which simplifies interaction while still enabling precise measurement of spatial memory. The level is intentionally designed without a time limit; once the participant is satisfied with the configuration, they terminate the level by activating a physical button in the room, which triggers the transition to the next scene or the end of the session.

Overall, the progression from guided navigation (Level 1) to unguided, timed, and perturbed navigation (Levels 2–5), culminating in a map-based task (Level 6), is designed to gradually increase task demands while promoting allocentric map learning and enabling a differentiated assessment of spatial strategies across the different conditions.

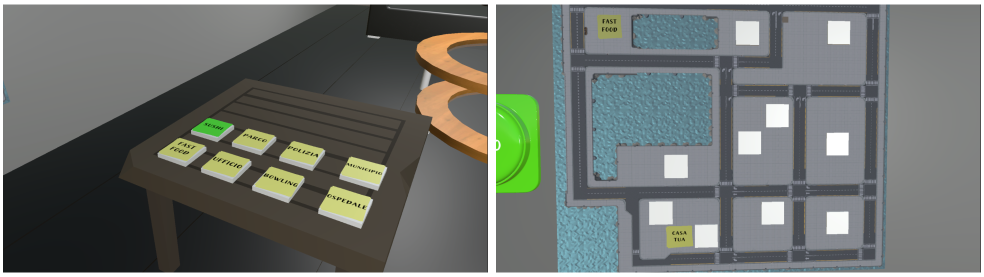


Fig. 6: Landmark placement task. On the left the table with the various post-its to be placed; on the right, the map with the empty placement positions.

3.1.4 Logging Data: Movement Heatmap and Landmark Observation

The data collection layer of *City Walk* was designed to capture behavioral and gaze information during navigation while remaining modular and decoupled from the core game logic. Three independent components are responsible for recording: the route taken by the player and the corresponding movement heatmap, performance in the city map scene, and eye-tracking metrics for landmark observation.

Movement Heatmap – During the navigation levels, the position of the player avatar is sampled continuously and accumulated into a two-dimensional heatmap defined over the city. The heatmap is implemented as an imaginary grid placed over the entire environment, with all cells having the same size. For each cell, the logger stores the number of seconds that the player spends within its boundaries. This representation makes it possible to reconstruct the path taken, estimate the total distance traveled in each level, and quantify local indecision as prolonged dwell time in specific cells (e.g., near intersections). Heatmaps from different levels can be aggregated to analyze which areas of the city were visited most frequently and which associative cues or landmarks were likely to have been encoded.

The grid definition is based on the dimensions of an invisible cube that spans the entire city. At runtime, the heatmap component reads the position and size of this cube and uses inspector-defined parameters for cell width and height to split the space into a matrix of cells. The player's position at each frame is then mapped to one of these cells, and the corresponding time counter is incremented. This solution automatically adapts to changes in the layout or extent of the city: when the terrain size or shape is modified, only the bounding cube needs to be adjusted, and the grid is recomputed accordingly, avoiding manual reconfiguration of the heatmap structure.

Landmark Observation – Eye tracking is used to quantify how participants visually engage with landmarks during navigation in the E-IVR condition. At each frame update, a ray is cast from the gaze origin along the gaze direction, and the hit object (if any) is checked for a specific landmark tag. For each landmark and for each level, two main variables are recorded: the total time the player spent looking at that landmark and the number of times the landmark was observed for at least a predefined number of consecutive seconds. We operationalized Conscious Observations

(COs) of landmarks using a conservative dwell-time criterion. Inspired by previous research (Hooge et al. 2025; Enders et al. 2021), we counted a conscious observation whenever the gaze ray remained on the same landmark for at least 2s. Shorter looks were treated as incidental glances and contributed only to total dwell time but not to the count of COs.

3.2 Experimental Design

We conducted a pilot study to obtain a preliminary evaluation of *City Walk* and to assess the feasibility of the experimental protocol. The study compared two conditions (DVR and E-IVR) in a between-subjects design, with participants randomly assigned to one of two groups. The study has been conducted in a controlled laboratory setting, and has been approved by the ethical committee of our university (protocol RM-2025-976).

The experimental procedure followed the same sequence for all participants and was completed in a single session. After providing informed consent and completing a brief pre-experience questionnaire, participants received instructions on the goal of the game and on the controls specific to their assigned condition. They then performed the full *City Walk* sequence used in the study configuration: an initial acclimatization level with guided navigation through the city, followed by four navigation levels with varying start–goal pairs and time constraints, and finally the map-based landmark placement level, in which eight landmarks had to be located on a schematic map of the city. Throughout gameplay, the system automatically logged in-game behavior and, in the E-IVR condition, eye-tracking data.

At the end of the game session, participants completed a set of post-experience questionnaires outside the game environment, assessing perceived cognitive load, user experience, cybersickness, system usability, and providing qualitative feedback on the application. All measures and analyses reported in this work should therefore be interpreted as preliminary, reflecting the pilot nature of the study and its primary goals of testing the apparatus, validating the logging pipeline, and obtaining initial trends on the effects of immersion and locomotion on spatial learning.

3.2.1 Experimental Measures

Description of all the experimental measures collected for the experiment: log from the application + questionnaires for the evaluation of the application.

During the pilot study, we collected a combination of objective in-game measures and subjective self-report measures, with the aim of characterizing navigation performance, spatial learning, landmark use, usability, cognitive workload, and cybersickness in the two experimental conditions.

In-game navigation performance – During each navigation level, the system logged the player's position over time on an imaginary grid superimposed on the city layout, yielding a heatmap in which each cell stores the number of seconds spent in that area. This representation was used to reconstruct the path followed on each level and to compute basic navigation performance indices, in particular:

- Total path length per level, obtained from the sequence of visited cells and the underlying spatial layout, providing an estimate of the distance traveled between the start and goal locations.
- Level completion time, defined as the time elapsed between the activation of the level and the arrival at the goal (or time-out).
- Delta distance, defined as the difference between the distance actually traveled and the optimal distance between start and goal for that level. This measure quantifies the degree of deviation from the optimal route and was later aggregated across levels to obtain an overall index of route efficiency for each participant.

These measures were extracted for all navigation levels used in the pilot and served to compare route efficiency and temporal performance between the two experimental conditions.

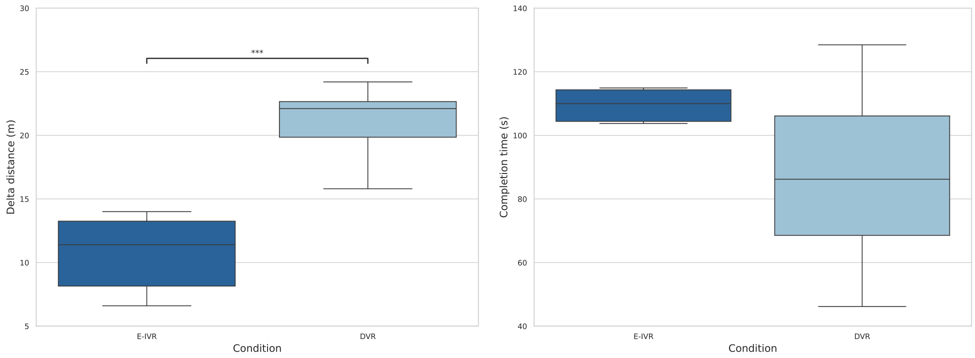
Landmark placement – At the end of the navigation phase, participants completed a city map scene in which they were asked to place notes representing the landmarks encountered in the city onto a 2D map viewed from above. For each note, the system recorded: whether it was placed on the map or left on the table, whether it was placed in the correct placeholder, and the spatial error in case of incorrect placement, computed as the distance between the correct placeholder and the current position of the note on the map. From these raw data, two aggregated indices were derived for each participant:

- Accuracy score, computed from the proportion of correctly placed landmarks.
- Total map error, defined as the sum of the individual spatial errors over all notes placed on the map.

Landmark observation – In the E-IVR condition, the eye-tracking component recorded participants' observation of predefined landmarks during navigation. For each landmark and for each level, the following measures were computed:

- Total observation time, defined as the cumulative number of seconds during which the gaze ray intersected the landmark.
- Number of COs, obtained by counting the number of times the landmark was observed for at least a predefined threshold of consecutive seconds.

Questionnaire – Outside the game, participants were asked to fill out two questionnaires: one pre-experiment to collect socio-demographic information and one post-experience to assess workload, user experience, cybersickness, and usability. Workload was assessed through the NASA Task Load Index (TLX) (Hart and Staveland 1988), composed by six items ranging from 0 (no effort) to 20 (maximum effort). User experience was evaluated by means of the User Experience Questionnaire (UEQ), by Laugwitz et al. (2008), comprising 26 bipolar adjective pairs rated on a 1–7 scale. Usability was assessed through the System Usability Scale (SUS) (Brooke 1995), which included 10 statements rated from 1 to 5. Finally, cybersickness was measured through the Virtual Reality Sickness Questionnaire (VRSQ), by Kim et al. (2018). Finally, we asked our participants to describe the application with three words, to grasp qualitative feedback on the project.



(a) Delta distance.

(b) Completion time.

Fig. 7: Delta distance and completion time results, divided by condition.

4 Experimental Results

The pilot study involved $N = 20$ participants aged between 18 and 35 years ($m = 24.4$, $std = 2.56$). Participants were randomly assigned to one of two experimental conditions, DVR ($m_{age} = 25$, $std = 2.75$) and E-IVR ($m_{age} = 23.8$, $std = 2.35$), with $n = 10$ per group.

4.1 In-App Performances

Distance and Completion Time – For each participant, we computed the *average delta distance* across the navigation levels, defined as the difference between the traveled distance and the corresponding optimal distance (averaged across levels). Normality checks did not reject the normality assumption (Shapiro–Wilk $p > 0.20$ in both groups), and homogeneity of variance was satisfied, hence we performed an independent-samples t-test, which indicated a statistically significant difference between conditions, with lower delta distances in E-IVR ($t(18) = 7.86$, $p < 0.001$, Cohen's $d = 3.52$), suggesting more efficient route execution in the immersive condition (Figure 7a).

In addition, we analyzed the *average completion time* across navigation levels. Normality checks again showed a normal distribution of both groups, and homogeneity of variance was satisfied. The between-condition difference did not reach statistical significance ($t(18) = -2.04$, $p = 0.056$; Cohen's $d = -0.91$), but showed a trend toward longer completion times in E-IVR (Figure 7b). Notably, the effect size was large in magnitude, suggesting that this pilot sample may have been underpowered to detect the observed difference and that a larger sample could clarify whether the apparent slowing in E-IVR reflects a reliable locomotion-related cost. The aggregated results are reported in Table 1.

We then evaluated the two metrics – delta distance and completion time – per level. Normality checks indicated non-normal distributions for delta distance at the level scale in at least one group (depending on the level); therefore, between-condition

Condition	**Delta distance**		**Time**	
	Mean	*SD*	*Mean*	*SD*
DVR	21.76	3.39	87.48	26.21
E-IVR	10.80	2.82	113.04	29.67

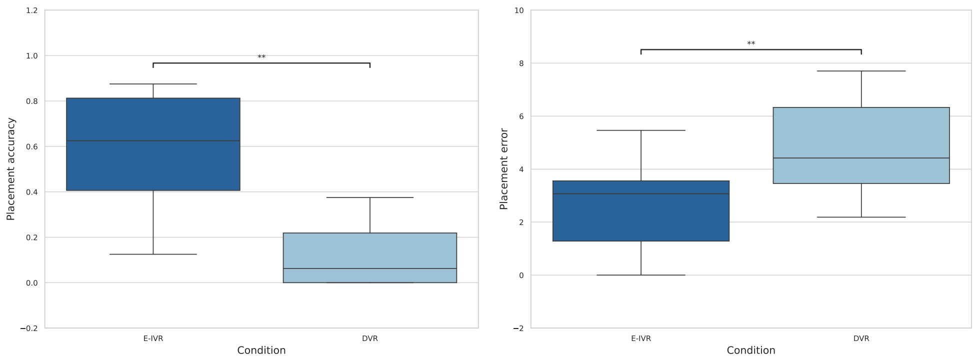
Table 1: Descriptive statistics (mean and standard deviation) for aggregated navigation measures by condition.

comparisons were conducted using non-parametric Mann–Whitney U tests. Delta distance was significantly lower in E-IVR than DVR in Level 2 ($U = 89.5$, $p = 0.002$), Level 3 ($U = 85.0$, $p = 0.007$), and Level 4 ($U = 86.5$, $p = 0.005$), while no significant between-condition differences emerged for Level 1 ($U = 71.0$, $p = 0.123$) and Level 5 ($U = 69.0$, $p = 0.165$).

Completion time was analyzed per level using independent-samples t-tests when normality assumptions were not rejected (Levels 1, 3, and 5) and Mann–Whitney U tests otherwise (Levels 2 and 4). Time did not significantly differ between conditions in Level 1 ($t(18) = -1.80$, $p = 0.089$, $d = -0.81$), Level 2 ($U = 44.0$, $p = 0.684$), or Level 3 ($t(18) = -1.21$, $p = 0.244$, $d = -0.54$). Conversely, E-IVR showed significantly longer completion times in Level 4 ($U = 14.0$, $p = 0.005$) and Level 5 ($t(18) = -3.51$, $p = 0.0025$, $d = -1.57$).

Landmark Placement – Spatial learning was assessed through the final city-map task using two complementary indicators: an *accuracy* measure, defined as the number of landmarks placed in the correct placeholder (regardless of the residual distance within the map), and the *total placement error*, defined as the sum of landmark-specific placement errors. For accuracy, DVR participants correctly placed fewer landmarks ($m = 1.2$, $sd = 1.69$) than E-IVR participants ($m = 4.6$, $sd = 2.12$). Normality was rejected for the overall sample and for the DVR group (Shapiro–Wilk $p < 0.05$); therefore, we compared conditions using a Mann–Whitney U test. Accuracy was significantly higher in E-IVR than DVR ($U = 10.5$, $p = 0.002$), indicating a higher number of correctly placed landmarks in the immersive condition (Figure 8a).

Spatial learning was also assessed through the final city-map task by computing, for each participant, the *map placement accuracy*, defined as the proportion of landmarks placed in the correct placeholder. Descriptively, DVR participants showed lower accuracy ($m = 0.15$, $sd = 0.21$) than E-IVR participants ($m = 0.58$, $sd = 0.26$). Normality checks indicated that the accuracy distribution violated the normality assumption overall and in the DVR group (Shapiro–Wilk $p = 0.019$; $p = 0.007$ for DVR). Accordingly, we used a non-parametric Mann-Whitney U test, revealing a statistically significant difference between conditions ($U = 10.5$, $p = 0.002$), with higher map-placement accuracy in E-IVR, indicating better landmark–location memory in the immersive embodied condition (Figure 8b). Overall descriptives are reported in Table 2.



(a) Placement accuracy.

(b) Total placement error.

Fig. 8: Placement accuracy and total placement error results, divided by condition.

Condition	**Accuracy (% correct)**		**Map error (total)**	
	Mean	*SD*	*Mean*	*SD*
DVR	0.15	0.21	4.86	1.91
E-IVR	0.58	0.26	2.64	1.66

Table 2: Landmark placement performances by condition.

Eye Tracking – Eye-tracking data were collected only in the E-IVR condition; therefore, results are reported descriptively and through within-group association analyses. For each landmark building, we computed the aggregated observation time and the aggregated number of times the landmark was watched, and related these measures to the corresponding landmark-specific placement error in the final city-map task. Because the goal was to capture monotonic associations in a small pilot sample, we used Spearman correlations (one correlation per landmark and per gaze metric).

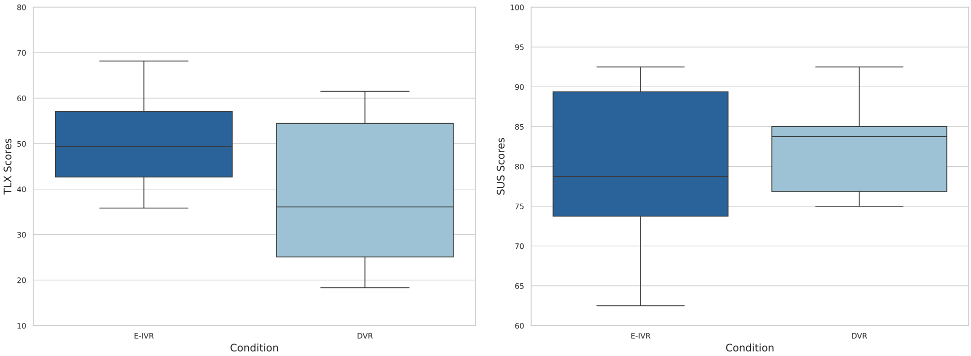
Overall, most landmarks did not show a reliable association between gaze behavior during navigation and subsequent placement accuracy. The only statistically significant correlations emerged for the *Bowling* landmark, where longer observation time was associated with *higher* map-placement error ($\rho = 0.768$, $p = 0.009$). No other landmark showed significant correlations, with the overall results reported in Table 3.

4.2 Questionnaire Results

Workload – Perceived workload was assessed using the NASA Task Load Index (TLX), summarized as the overall TLX score. Descriptively, workload was lower in DVR ($m = 38.77$, $sd = 16.81$) than in E-IVR ($m = 50.90$, $sd = 11.39$). Normality checks did not reject the normality assumption for either group (Shapiro–Wilk $p > 0.10$). However, homogeneity of variance was not satisfied (Levene's test $p = 0.04$);

Landmark	ρ **(observation time vs. error)**	p
Bowling	0.768	0.009
FastFood	-0.256	0.474
Hospital	0.072	0.844
Municiple	-0.112	0.758
Office	0.078	0.831
Park	0.350	0.321
PoliceStation	-0.165	0.650
Sushi	-0.675	0.066

Table 3: Spearman correlations between landmark dwell time and map-placement error (E-IVR).



(a) TLX results, divided by condition.

(b) SUS results, divided by condition.

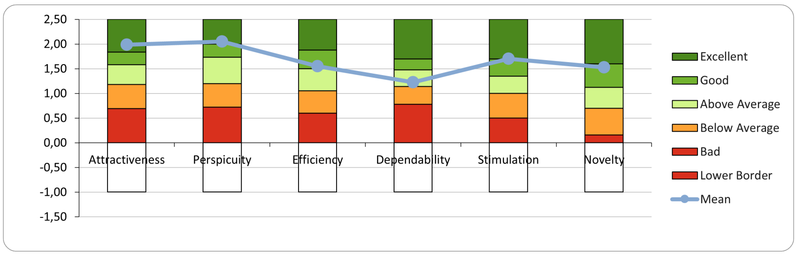
Fig. 9: TLX and SUS results.

therefore, we used Welch's t-test. The between-condition difference did not reach conventional significance ($t = 1.89$, $p = 0.077$, Cohen's $d = 0.845$), but showed a trend toward higher perceived workload in E-IVR (Figure 9a).

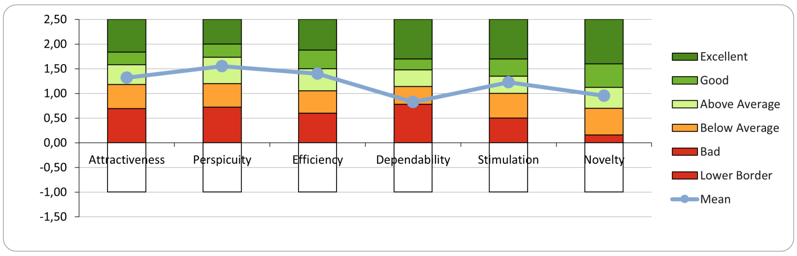
Usability – Descriptively, SUS scores were comparable across conditions, with DVR showing $m = 78.25$ ($sd = 13.95$) and E-IVR showing $m = 77.50$ ($sd = 15.59$). Following the grade-scale by Bangor et al. (2009), both conditions can be rated as "good", indicating a good level of usability with room for improvements.

Normality was rejected for the overall sample and for the DVR group (Shapiro–Wilk $p < 0.05$), therefore we compared conditions using a Mann–Whitney U test. No statistically significant difference emerged between DVR and E-IVR ($U = 50.5$, $p = 1.0$), suggesting similar perceived usability across the two conditions (Figure 9b).

User Experience – User experience was assessed through the UEQ, reporting the six standard scales (Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty) on the transformed $[-3, +3]$ range, where positive values indicate a positive evaluation. Across conditions, all scale means were above zero, suggesting an



(a) UEQ benchmark for the E-IVR condition.



(b) UEQ benchmark for the DVR condition.

Fig. 10: UEQ benchmarks, generated through the tool provided by Schrepp et al. (2017).

overall positive experience in both DVR and E-IVR (**report data in table**). Between-condition comparisons (two-sample t-tests assuming unequal variances) did not reveal statistically significant differences for any UEQ scale (all $p \geq 0.187$; Attractiveness $p = 0.187$, Perspicuity $p = 0.256$, Efficiency $p = 0.731$, Dependability $p = 0.189$, Stimulation $p = 0.466$, Novelty $p = 0.298$). Using the UEQ benchmark categorization (Schrepp et al. 2017), E-IVR was classified as *Excellent* for Attractiveness and Perspicuity and *Good* for Efficiency, Stimulation, and Novelty, while Dependability was *Above Average* (Figure 10a). In DVR, most scales were *Above Average*, whereas Dependability was categorized as *Below Average* (Figure 10b).

Cybersickness – Cybersickness was assessed using the Virtual Reality Sickness Questionnaire (VRSQ), considering the oculomotor subscale, the disorientation subscale, and the total score. Descriptively, oculomotor symptoms were similar across conditions (DVR: $m = 2.4$, $sd = 1.84$; E-IVR: $m = 2.8$, $sd = 2.66$). Disorientation symptoms were also comparable (DVR: $m = 2.1$, $sd = 2.73$; E-IVR: $m = 1.7$, $sd = 2.06$). Finally, total VRSQ scores were identical in mean across conditions (DVR: $m = 4.5$, $sd = 3.69$; E-IVR: $m = 4.5$, $sd = 4.03$).

Normality was rejected for the overall distributions (and for at least one group in each scale); therefore, we used Mann–Whitney U tests to compare conditions. No statistically significant differences emerged between DVR and E-IVR for the oculomotor subscale ($U = 48.5$, $p = 0.971$), disorientation subscale ($U = 52.0$, $p = 0.912$), or total score ($U = 51.5$, $p = 0.971$), suggesting comparable self-reported cybersickness across the two conditions.

(a) DVR

(b) E-IVR

Fig. 11: Wordclouds of adjectives used to describe the application, split by condition. Colors indicate polarity (green: positive; yellow: neutral; red: negative), while dimension indicate the frequency of the words in the dataset.

Qualitative – Figure 11 summarizes the adjectives most frequently used to describe the application in each condition. In both DVR and E-IVR, participants predominantly selected positive descriptors (green), with *innovative* and *intuitive* emerging as the most recurring terms. Negative descriptors (red) were comparatively less frequent and mainly related to effort and discomfort (e.g., *tiring*, *uncomfortable*). Neutral descriptors (yellow), such as *challenging* and *demanding*, appeared in both conditions and suggest that some participants perceived the task as cognitively engaging rather than uniformly easy.

5 Discussion

We presented a pilot to validate the feasibility of the *City Walk* protocol and provide preliminary evidence on whether higher immersion and more natural locomotion (E-IVR) offer measurable advantages for spatial learning compared to a desktop baseline (DVR). Accordingly, all findings should be interpreted as initial trends that motivate confirmatory studies with larger samples and the intended target population.

5.1 Immersion and Locomotion: Implications for Navigation Efficiency and Spatial Learning

The aggregated navigation outcomes suggest a clear dissociation between route efficiency and execution speed; the level-based analyses clarify where this trade-off is expressed within *City Walk*'s progression. Participants in E-IVR showed a significantly lower delta distance than participants in DVR, indicating more efficient navigation and fewer detours or backtracking. Notably, this advantage emerged in the levels designed to probe allocentric learning and flexible route planning (Levels 2–4). By contrast, Level 5 – which evaluates transfer of acquired spatial knowledge from a novel start under a strict time limit – did not show significant differences, plausibly reflecting greater strategic variability and/or ceiling effects imposed by the time cap. From a spatial-learning perspective, this level-selective pattern is consistent with frameworks distinguishing egocentric stimulus–response route execution from allocentric, map-like representations: because Levels 2–4 require integrating inter-landmark relationships across changing start/goal configurations, they should preferentially benefit from allocentric encoding and flexible inference rather than fixed turn sequences (Klatzky 1998; Ekstrom et al. 2014; Iglói et al. 2009; Chan et al. 2012; Wolbers and Hegarty 2010). Moreover, embodied locomotion provides richer self-motion cues (e.g., vestibular and proprioceptive information) that support spatial updating and path integration; when combined with stable landmarks, these cues can reduce detours and improve route efficiency even if they do not minimize elapsed time (Taube et al. 2013; Ruddle and Lessels 2009; Diersch et al. 2019; Hejtmanek et al. 2020; Riecke et al. 2010). The absence of significant differences in Level 5 may therefore reflect the combination of a novel-start transfer demand and a hard time cap, which increases strategy heterogeneity and truncates path expression, making delta distance less sensitive to representational advantages under throughput-limited embodied locomotion (Wolbers and Hegarty 2010; Ekstrom et al. 2014; Ruddle and Lessels 2009; Riecke et al. 2010).

Conversely, completion time tended to be lower in DVR than in E-IVR; although this difference was not significant in the aggregated comparison, the large Cohen's d suggests the pilot may have been underpowered. The level-wise pattern further indicates that any time cost is concentrated in the most demanding stages, with E-IVR participants significantly slower in Levels 4 and 5. This pattern is consistent with a meaningful locomotion-drive cost: mouse–keyboard movement can enable rapid translation and turning even when the chosen route is suboptimal (Ruddle and Lessels 2009; Zhao et al. 2020; Lapointe et al. 2011), whereas treadmill-based locomotion imposes physical constraints (walking pace, turning mechanics, slower reorientation) that may slow traversal (Lewis et al. 2024; Soni and Lamontagne 2020; Pereira et al. 2025) – particularly in levels requiring detours and rapid re-planning – even when participants select more optimal paths.

Taken together, these results point to a trade-off: E-IVR appears to support better decision-making and route selection (efficiency), while DVR interactions may facilitate faster completion by allowing higher movement throughput. Given the aim of this project was to evaluate how immersion and embodied locomotion shape spatial navigation learning, delta distance is the key outcome, because it more directly captures the quality of navigation choices and the formation/use of spatial knowledge, whereas time is partially confounded by interface mechanics. Conversely, if the primary aim is to identify the fastest strategy to complete a level (e.g., performance optimization, speedrunning-like design goals, or minimizing time-on-task), an embodied locomotion setup like E-IVR may not be the optimal solution.

In the landmark identification task, E-IVR participants showed a clear advantage in survey-level spatial memory as indexed by map-placement accuracy (proportion of correctly placed landmarks). Because the task was designed with more placeholders than targets, performance is less likely to reflect simple elimination and more likely to reflect the quality of an internal representation of landmark locations and their relations. In line with spatial–cognition accounts distinguishing egocentric stimulus–response strategies from allocentric map–like representations, these findings suggest that E-IVR better supported encoding and/or retrieval of inter-landmark structure that can be expressed in a survey–format task (Klatzky 1998; Ekstrom et al. 2014; Wolbers and Hegarty 2010). Mechanistically, this improvement is compatible with the notion that embodied locomotion provides richer self-motion cues that can facilitate spatial updating and integration of landmark information during learning (Taube et al. 2013; Ruddle and Lessels 2009; Riecke et al. 2010). Overall, landmark placement results – coupled with delta distance – provide converging evidence that E-IVR enhances spatial knowledge acquisition when compared with a DVR setup.

To investigate potential mechanisms underlying the E-IVR advantage in delta distance and map performance, we examined whether landmark knowledge could be attributed to increased visual sampling of landmarks during navigation. Eye-tracking indices of visual engagement were only weakly related to participants' landmark knowledge, as expressed in the post-task map placement. In particular, for most landmarks – including the core beacons explicitly emphasized during navigation and several intermediate waypoints – there was no significant correlation between map-placement error and total landmark observation time. This pattern is consistent with the intended design in which core landmarks can be localized using the city's global structure (e.g., corner placement and coastline reference) and intermediate landmarks may be used procedurally as route cues without necessarily being integrated into a metrically precise allocentric representation; in such cases, overt fixation measures are not expected to be strong predictors of later 2D placement accuracy (Viaene et al. 2016; Segen et al. 2022; Piccardi et al. 2016). The only landmark showing a reliable association was the Bowling, indicating that participants who looked at this landmark more tended to place it less accurately. Rather than implying that attention harms encoding, this directionality is compatible with a compensatory interpretation: longer dwell times may indicate increased processing demands or difficulty (Horstmann et al. 2016; Rayner 1998) and, in way-finding contexts, extended fixations and repeated scanning/attention shifts are commonly observed under unfamiliarity or spatial ambiguity and can reflect cautious inspection or unsuccessful visual search rather than efficient spatial integration (Kwon et al. 2025; Viaene et al. 2016).

Overall, this multi-metric analysis provides a coherent account of how immersion and locomotion shape performance in *City Walk*: embodied locomotion appears to support the acquisition of spatial knowledge that generalizes beyond single-route execution, while time-based performance remains strongly influenced by interface-specific locomotion throughput. Methodologically, the combination of path-efficiency metrics, a survey-format map endpoint, and in-app logging helps disentangle learning-related gains from interaction costs, offering a practical template for evaluating VR navigation tools beyond completion time alone. The eye-tracking probe further suggests that spatial learning benefits are unlikely to be explained by a simple "more looking equals better memory" mechanism, motivating richer gaze features and complementary indices of uncertainty and replanning. These findings position *City Walk* as a promising platform for studying navigation learning under controlled manipulations of locomotion, and they motivate larger-sample validation with refined process measures.

5.2 User Experience, Usability and Cybersickness: Acceptability of the Approach

A complementary aim of this pilot was to evaluate *City Walk* from a user-centered perspective, focusing on perceived workload, overall user experience, usability, and collecting data on any possible sickness symptoms elicited by the application. Across instruments, the questionnaire pattern supports the acceptability of both implementations: participants reported generally positive experiences and comparable usability, and the E-IVR version did not produce higher self-reported cybersickness relative to the desktop baseline.

Perceived workload did not differ reliably between conditions. While descriptive patterns are consistent with the intuition (and prior VR work) that adding embodied locomotion can increase cognitive–motor demands (Bruder et al. 2015; Brument et al. 2025; Sarupuri et al. 2017) – particularly when physical walking and body-based turning are required rather than mouse–keyboard control – the present pilot provides no robust evidence that E-IVR is experienced as systematically more demanding. Given the small sample size and the pilot purpose of validating feasibility, the most conservative interpretation is that workload is broadly comparable at this stage, with any subtle differences likely requiring a larger sample (and the intended target population) to estimate precisely.

User experience and usability outcomes converge on a similarly encouraging conclusion. UEQ results indicate a positive perception of the application in both conditions and do not suggest that immersion or treadmill locomotion compromises perceived clarity, efficiency, or overall appeal. SUS responses likewise indicate that participants found the system usable, supporting the feasibility of deploying the protocol without major usability-related barriers. From a design standpoint, the UEQ benchmark pattern also usefully highlights where iteration is most valuable (e.g., strengthening perceived dependability and consistency in the desktop version through clearer feedback, state transitions, and error prevention), without implying that either version is unusable.

Cybersickness outcomes further support tolerability: VRSQ responses did not indicate an increase in sickness symptoms in E-IVR relative to DVR. This is a particularly important feasibility signal for any VR application intended for repeated sessions, where cybersickness is a known practical barrier to sustained use (Caserman et al. 2021; Clifton and Palmisano 2020). The qualitative adjective data help refine this interpretation by suggesting that negative impressions were more often framed in terms of effort and discomfort than classic sickness symptomatology (Figure 11), implying that design improvements should prioritize ergonomic comfort (e.g., harness fit, pacing, opportunities for rest) separately from sickness-mitigation strategies.

Taken together, the questionnaire evidence indicates that *City Walk* is broadly acceptable in both DVR and E-IVR, and that the added immersion and treadmill-based locomotion – while likely introducing additional physical effort by construction – does not appear to undermine usability or tolerability in this pilot. Future work should confirm these user-centered findings in larger samples and, critically, in the intended middle-aged/older adult population, where comfort, fatigue, and accessibility constraints are expected to play a larger role in overall acceptability.

5.3 Limitations

This project is not exempt from limitations. Firstly, the study was conducted with a small between-subjects sample ($N = 20$; $n = 10$ per condition), which limits statistical power and make results more susceptible to outliers. This constraint is especially relevant for landmark-level eye-tracking associations, where per-landmark data availability can further reduce effective sample size and inflate uncertainty. Importantly, the pilot nature of this work implies that larger confirmatory studies are already planned.

Although the longer-term goal of *City Walk* is to support spatial-navigation assessment/training in middle-aged and older adults, the pilot sample comprised young adults. As a result, the present findings do not support generalization to the intended target population, where age-related changes in spatial strategies, fatigue, balance, and cognitive–motor interference could substantially alter both usability and performance outcomes.

Landmark observation was operationalized using a conservative COs definition (continuous gaze-on-landmark duration $\geq 2s$). While defensible for capturing sustained engagement, this threshold likely reduced sensitivity in a wayfinding context, where meaningful sampling may occur via shorter fixations and repeated brief returns (particularly at intersections and during active scanning). As a result, the absence of robust landmark-wise associations should be interpreted cautiously; future work should adopt lower dwell thresholds and/or standard fixation parsing, and incorporate revisit-based metrics rather than relying on a single conservative criterion

A further limitation is that the embodied locomotion condition can introduce confounds related to pacing, physical effort, and fatigue. Longer traversal times may simultaneously increase exposure to the environment (potentially benefiting later spatial-memory measures) and reflect the motor/ergonomic demands of treadmill walking and turning rather than purely cognitive effects of immersion. Without independent measures of exertion and fatigue, it remains unclear whether time differences reflect increased immersion-related processing, increased physical load, or both. Future studies should explicitly dissociate cognitive workload from physical exertion, for instance through physiological measures.

6 Conclusions and Future Research Directions

This work introduced *City Walk*, a VR serious-game designed to elicit and measure spatial navigation learning within a controlled urban environment, while systematically manipulating immersion and locomotion. The protocol combines multi-level navigation with escalating demands (guided familiarization, unguided exploration, route reversal, obstacle-based re-planning, and transfer from a novel start) and culminates in a survey-format landmark placement task. This structure, together with in-app logging (movement heatmaps and eye-tracking–based landmark observation in the E-IVR build) and user-centered questionnaires, was intended to jointly assess navigation performance, spatial knowledge acquisition, and feasibility of the application as a training-oriented tool.

Across measures, the pilot results support a coherent interpretation: embodied immersive locomotion is associated with improved indicators of spatial learning and map-consistent knowledge, while time-based performance is more strongly influenced by locomotion throughput and interaction constraints. In particular, route-efficiency benefits were most evident in those levels designed to require flexible planning and re-planning, and the map-based landmark placement endpoint provided converging evidence that the learned representation in E-IVR was more consistent with a survey-level spatial model. At the same time, longer completion times in E-IVR highlight a practical trade-off for embodied locomotion, and exploratory eye-tracking analyses suggest that spatial-learning benefits are not reducible to a simple “more looking equals better memory” mechanism. From an application perspective, the user-centered measures indicate that the platform is usable and tolerable in this pilot setting, supporting feasibility for further validation.

Future work will refine gaze metrics by reducing the dwell threshold and incorporating revisit-based indices; evaluate the intended middle-aged/older adult population with explicit assessment of accessibility constraints (training duration, fatigue, balance demands) and cognitive–motor interference; and complement subjective workload with physiological measures to better distinguish cognitive workload from physical exertion and to interpret the observed efficiency–speed trade-off.

Statements and Declarations. This work was supported by the MUSA – Multilayered Urban Sustainability Action project, funded by the European Union – NextGenerationEU initiative.

The authors have no relevant financial or non-financial interests to disclose.

References

Abd-Alrazaq, A., Alajlani, M., Alhuwail, D., Toro, C.T., Giannicchi, A., Ahmed, A., Makhlouf, A., Househ, M.: The effectiveness and safety of serious games for improving cognitive abilities among elderly people with cognitive impairment: systematic review and meta-analysis. JMIR serious games **10**(1) (2022) https://doi.org/10.2196/34592

Ackerman, P.L., Kanfer, R., Calderwood, C.: Use it or lose it? wii brain exercise practice and reading for domain knowledge. Psychology and Aging **25**(4), 753–766 (2010) https://doi.org/10.1037/a0019277

Buttussi, F., Chittaro, L.: Locomotion in place in virtual reality: A comparative evaluation of joystick, teleport, and leaning. IEEE Transactions on Visualization and Computer Graphics **27**(1), 125–136 (2021) https://doi.org/10.1109/TVCG.2019.2928304

Boffi, P., Clerici, M., Mauri, G., Meneghin, G., Gallace, A., Lanzi, P.L.: Comparing traditional and haptic controllers for fine-movement tasks in virtual reality. In: 2024 IEEE 12th International Conference on Serious Games and Applications for Health (SeGAH), pp. 1–8 (2024). https://doi.org/10.1109/SeGAH61285.2024.10639530

Brument, H., Gerdenich, M., Kaufmann, H.: Higher cognitive load increases unintended positional drift during navigation in virtual reality. In: Proceedings of the

2025 31st ACM Symposium on Virtual Reality Software and Technology. Association for Computing Machinery, ??? (2025). https://doi.org/10.1145/3756884.3765968

Bangor, A., Kortum, P., Miller, J.: Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. J. Usability Studies **4**(3), 114–123 (2009). [Online]. Available here

Bruder, G., Lubas, P., Steinicke, F.: Cognitive resource demands of redirected walking. IEEE Transactions on Visualization and Computer Graphics **21**(4), 539–544 (2015) https://doi.org/10.1109/TVCG.2015.2391864

Brooke, J.: Sus: A quick and dirty usability scale. Usability Eval. Ind. **189** (1995). [Online]. Available **here**

Bliss, J.P., Tidwell, P.D., Guest, M.A.: The effectiveness of virtual reality for administering spatial navigation training to firefighters. Presence: Teleoperators and Virtual Environments **6**(1), 73–86 (1997) https://doi.org/10.1162/pres.1997.6.1.73

Checa, D., Bustillo, A.: A review of immersive virtual reality serious games to enhance learning and training. Multimedia Tools and Applications **79**, 5501–5527 (2020) https://doi.org/10.1007/s11042-019-08348-9

Chan, E., Baumann, O., Bellgrove, M.A., Mattingley, J.B.: From objects to landmarks: The function of visual location information in spatial navigation. Frontiers in Psychology **3** (2012) https://doi.org/10.3389/fpsyg.2012.00304

Callahan-Flintoft, C., Barentine, C., Touryan, J., Ries, A.J.: A case for studying naturalistic eye and head movements in virtual environments. Frontiers in Psychology **12**, 650693 (2021) https://doi.org/10.3389/fpsyg.2021.650693

Caserman, P., Garcia-Agundez, A., Gámez Zerban, A., Göbel, S.: Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook. Virtual Reality **25**, 1153–1170 (2021) https://doi.org/10.1007/s10055-021-00513-6

Cipresso, P., Giglioli, I., Raya, M., Riva, G.: The past, present, and future of virtual and augmented reality research: A network and content analysis of articles published in frontiers in psychology. Frontiers in Psychology **9**, 2086 (2018) https://doi.org/10.3389/fpsyg.2018.02086

Chakraborty, S., Kane, A., Gagnon, H.C., McNamara, T.P., Bodenheimer, B.: Comparative effectiveness of an omnidirectional treadmill versus natural walking for navigating in virtual environments. In: ACM Symposium on Applied Perception 2024 (SAP '24) (2024). https://doi.org/10.1145/3675231.3675243

Chang, E., Kim, H.T., Yoo, B.: Virtual Reality Sickness: A Review of Causes and

Measurements. International Journal of Human-Computer Interaction, 1658–1682 (2020) https://doi.org/10.1080/10447318.2020.1778351

Capriotti, A., Moret, S., Del Bello, E., Federici, A., Lucertini, F.: Virtual reality: A new frontier of physical rehabilitation. Sensors **25**(10) (2025) https://doi.org/10.3390/s25103080

Clifton, J., Palmisano, S.: Effects of Steering Locomotion and Teleporting on Cybersickness and Presence in HMD-based Virtual Reality. Virtual Reality **24**, 453–468 (2020) https://doi.org/10.1007/s10055-019-00407-8

Cogné, M., Taillade, M., N'Kaoua, B., Tarruella, A., Klinger, E., Larrue, F., Sauzéon, H., Joseph, P.-A., Sorita, E.: The contribution of virtual reality to the diagnosis of spatial navigation disorders and to the study of the role of navigational aids: A systematic literature review. Annals of Physical and Rehabilitation Medicine **60**(3), 164–176 (2017) https://doi.org/10.1016/j.rehab.2015.12.004

Clemenson, G.D., Wang, L., Mao, Z., Stark, S.M., Stark, C.E.L.: Exploring the spatial relationships between real and virtual experiences: What transfers and what doesn't. Frontiers in Virtual Reality **1**, 572122 (2020) https://doi.org/10.3389/frvir.2020.572122

Di Natale, A.F., Repetto, C., Riva, G., Villani, D.: Immersive virtual reality in k-12 and higher education: A 10-year systematic review of empirical research. British Journal of Educational Technology **51**(6), 2006–2033 (2020) https://doi.org/10.1111/bjet.13030

Diersch, N., Wolbers, T., Jundi, B., Kelber, A., Webb, B.: The potential of virtual reality for spatial navigation research across the adult lifespan. Journal of Experimental Biology **222**, 187252 (2019) https://doi.org/10.1242/jeb.187252

Ekstrom, A.D., Arnold, A.E.G.F., Iaria, G.: A critical review of the allocentric spatial representation and its neural underpinnings: toward a network-based perspective. Frontiers in Human Neuroscience **8**, 803 (2014) https://doi.org/10.3389/fnhum.2014.00803

Ellis, R.: 1. Implicit and Explicit Learning, Knowledge and Instruction, pp. 3–26. Multilingual Matters, Bristol, Blue Ridge Summit (2009). https://doi.org/10.21832/9781847691767-003

Enders, L.R., Smith, R.J., Gordon, S.M., Ries, A.J., Touryan, J.: Gaze behavior during navigation and visual search of an open-world virtual environment. Frontiers in Psychology **12**, 681042 (2021) https://doi.org/10.3389/fpsyg.2021.681042

Fusco, A., Tieri, G.: Challenges and perspectives for clinical applications of immersive and non-immersive virtual reality. Journal of Clinical Medicine **11**(15) (2022) https://doi.org/10.3390/jcm11154540

Gröhn, M., Lokki, T., Takala, T.: Comparison of auditory, visual, and audiovisual navigation in a 3d space. ACM Trans. Appl. Percept. **2**(4), 564–570 (2005) https://doi.org/10.1145/1101530.1101558

Gabes, M., Mühlberger, A.: Driving simulation: the effects of interactivity and presentation setting. Frontiers in Virtual Reality **5** (2024) https://doi.org/10.3389/frvir.2024.1484739

Gutiérrez, F., Pierce, J., Vergara, V., Coulter, R., Saland, L., Caudell, T., Goldsmith, T., Alverson, D.: The effect of degree of immersion upon learning performance in virtual reality simulations for medical education. Studies in health technology and informatics **125**, 155–60 (2007) https://doi.org/10.1097/00042871-200701010-00099

Hooks, K., Ferguson, W., Morillo, P., Cruz-Neira, C.: Evaluating the user experience of omnidirectional VR walking simulators. Entertainment Computing **34**, 100352 (2020) https://doi.org/10.1016/j.entcom.2020.100352

Horstmann, G., Herwig, A., Becker, S.I.: Distractor dwelling, skipping, and revisiting determine target absent performance in difficult visual search. Frontiers in Psychology **Volume 7 - 2016** (2016) https://doi.org/10.3389/fpsyg.2016.01152

Hořejší, P., Lochmannová, A., Jezl, V., Dvořák, M.: Virtual reality locomotion methods differentially affect spatial orientation and cybersickness during maze navigation. Scientific Reports **15**, 26255 (2025) https://doi.org/10.1038/s41598-025-12143-y

Hooge, I.T.C., Nuthmann, A., Nyström, M., Niehorster, D.C., Holleman, G.A., Andersson, R., Hessels, R.S.: The fundamentals of eye tracking part 2: From research question to operationalization. Behav Res **57**(73) (2025) https://doi.org/10.3758/s13428-024-02590-2

Hart, S.G., Staveland, L.E.: Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: Human Mental Workload vol. 52, pp. 139–183 (1988). https://doi.org/10.1016/S0166-4115(08)62386-9

Hejtmanek, L., Starrett, M., Ferrer, E., Ekstrom, A.D.: How much of what we learn in virtual reality transfers to real-world navigation? Multisensory Research **33**(4-5), 479–503 (2020) https://doi.org/10.1163/22134808-20201445

Iglói, K., Zaoui, M., Berthoz, A., Rondi-Reig, L.: Sequential egocentric strategy is acquired as early as allocentric strategy: Parallel acquisition of these two navigation strategies. Hippocampus **19**(12), 1199–1211 (2009) https://doi.org/10.1002/hipo.20595

Jonson, M., Avramescu, S., Chen, D., Alam, F.: The role of virtual reality in screening, diagnosing, and rehabilitating spatial memory deficits. Frontiers in Human

Neuroscience **15**, 628818 (2021) https://doi.org/10.3389/fnhum.2021.628818

Jensen, L., Konradsen, F.: A review of the use of virtual reality head-mounted displays in education and training. Educ Inf Technol **23**, 1515–1529 (2018) https://doi.org/10.1007/s10639-017-9676-0

Jabbari, Y., Kenney, D.M., Mohrenshildt, M., Shedden, J.M.: Testing landmark-specific effects on route navigation in an ecologically valid setting: a simulated driving study. Cogn. Research **7**(22) (2022) https://doi.org/10.1186/s41235-022-00374-w

König, S.U., Clay, V., Nolte, D., Duesberg, L., Kuske, N., Peter, K.: Learning of spatial properties of a large-scale virtual city with an interactive map. Frontiers in Human Neuroscience **13** (2019) https://doi.org/10.3389/fnhum.2019.00240

Kelly, D.M., Gibson, B.M.: Spatial navigation: Spatial learning in real and virtual environments. Comparative Cognition & Behavior Reviews **2** (2007) https://doi.org/10.3819/ccbr.2008.20007

Klatzky, R.L.: In: Freksa, C., Habel, C., Wender, K.F. (eds.) Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections, pp. 1–17. Springer, Berlin, Heidelberg (1998). https://doi.org/10.1007/3-540-69342-4_1

Kim, H.K., Park, J., Choi, Y., Choe, M.: Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. Applied Ergonomics **69**, 66–73 (2018) https://doi.org/10.1016/j.apergo.2017.12.016

Kwon, J., Schmidt, A., Luo, C., Jun, E., Martinez, K.: Visualizing spatial cognition for wayfinding design: Examining gaze behaviors using mobile eye tracking in counseling service settings. ISPRS International Journal of Geo-Information **14**(10) (2025) https://doi.org/10.3390/ijgi14100406

Kober, S.E., Wood, G., Hofer, D., Kreuzig, W., Kiefer, M., Neuper, C.: Virtual reality in neurologic rehabilitation of spatial disorientation. Journal of neuroengineering and rehabilitation **10**, 1–13 (2013) https://doi.org/10.1186/1743-0003-10-17

Lackner, J.: Motion sickness: more than nausea and vomiting. Experimental Brain Research **232**(8), 2493–2510 (2014) https://doi.org/10.1007/s00221-014-4008-8

Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: HCI and Usability for Education and Work, pp. 63–76 (2008). https://doi.org/10.1007/978-3-540-89350-9_6

Lampropoulos, G., Kinshuk: Virtual reality and gamification in education: a systematic review. Educational Technology Research and Development **72**, 1691–1785 (2024) https://doi.org/10.1007/s11423-024-10351-3

Lee, J., Kim, M., Kim, J.: A study on immersion and vr sickness in walking interaction for immersive virtual reality applications. Symmetry **9**(5), 78 (2017) https://doi.org/10.3390/sym9050078

Lapointe, J.-F., Savard, P., Vinson, N.G.: A comparative study of four input devices for desktop virtual walkthroughs. Computers in Human Behavior **27**(6), 2186–2191 (2011) https://doi.org/10.1016/j.chb.2011.06.014

Lewis, M.M., Waltz, C., Scelina, L., Scelina, K., Owen, K.M., Hastilow, K., Zimmerman, E.M., Rosenfeldt, A.B., Koop, M.M., Alberts, J.L.: Gait patterns during overground and virtual omnidirectional treadmill walking. Journal of NeuroEngineering and Rehabilitation **21**(29) (2024) https://doi.org/10.1186/s12984-023-01286-6

Moreno-Arjonilla, J., López-Ruiz, A., Jiménez-Pérez, J.R., Callejas-Aguilera, J.E., Jurado, J.M.: Eye-tracking on virtual reality: a survey. Virtual Reality **28**(38) (2024) https://doi.org/10.1007/s10055-023-00903-y

Michael, D.R., Chen, S.: Serious Games: Games that Educate, Train, and Inform. Thomson Course Technology PTR, Boston, MA (2006)

Mortara, M., Catalano, C.E., Bellotti, F., Fiucci, G., Houry-Panchetti, M., Petridis, P.: Learning cultural heritage by serious games. Journal of Cultural Heritage **15**(3), 318–325 (2014) https://doi.org/10.1016/j.culher.2013.04.004

McLaren-Gradinaru, M., Burles, F., Dhillon, I., Leonidas David, A., Umiltà, A., Hannah, J., Dolhan, K., Iaria, G.: A novel training program to improve human spatial orientation: Preliminary findings. Frontiers in Human Neuroscience **14** (2020) https://doi.org/10.3389/fnhum.2020.00005

Monteiro, P., Peixoto, B., Gonçalves, G., Coelho, H., Barbosa, L., Melo, M., Bessa, M.: Beyond the hands: Evaluating the usability of hands-free methods and controllers for menu selection during an immersive VR experience. International Journal of Human–Computer Interaction, 1–38 (2025) https://doi.org/10.1080/10447318.2025.2553776

Oei, A.C., Patterson, M.D.: Enhancing cognition with video games: A multiple game training study. PLoS ONE **8**(3) (2013) https://doi.org/10.1371/journal.pone.0058546

Piccardi, L., De Luca, M., Nori, R., Palermo, L., Iachini, F., Guariglia, C.: Navigational style influences eye movement pattern during exploration and learning of an environmental map. Frontiers in Behavioral Neuroscience **10** (2016) https://doi.org/10.3389/fnbeh.2016.00140

Park, H., Dixit, M.K., Pariafsai, F.: Towards personally relevant navigation: The differential effects of cognitive style and map orientation on spatial knowledge development. Applied Sciences **14**(10) (2024) https://doi.org/10.3390/app14104012

Pereira, T.V., Gourdou, P., Sangani, S., Blanchette, A.K., Archambault, P.S., Lamontagne, A.: Turning strategies while walking on an omnidirectional treadmill with virtual reality. IEEE Transactions on Neural Systems and Rehabilitation Engineering **33**, 4523–4533 (2025) https://doi.org/10.1109/TNSRE.2025.3630093

Perdomo, B., Pérez, C.: Research on serious games: Bibliometric study of current status and future trends. TEM Journal, 805–815 (2025) https://doi.org/10.18421/TEM141-71

Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological Bulletin **124**(3), 372–422 (1998) https://doi.org/10.1037/0033-2909.124.3.372

Riecke, B.E., Bodenheimer, B., McNamara, T.P., Williams, B., Peng, P., Feuereissen, D.: Do we need to walk for effective virtual reality navigation? physical rotations alone may suffice. In: Hölscher, C., Shipley, T.F., Olivetti Belardinelli, M., Bateman, J.A., Newcombe, N.S. (eds.) Spatial Cognition VII. Lecture Notes in Artificial Intelligence, vol. 6222, pp. 234–247. Springer, ??? (2010). https://doi.org/10.1007/978-3-642-14749-4_21

Ruddle, R.A., Lessels, S.: The benefits of using a walking interface to navigate virtual environments. ACM Transactions on Computer-Human Interaction **16**(1), 1–18 (2009) https://doi.org/10.1145/1502800.1502805

Radianti, J., Majchrzak, T.A., Fromm, J., Wohlgenannt, I.: A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. Computers & Education **147**, 103778 (2020) https://doi.org/10.1016/j.compedu.2019.103778

Robinson, D.A.: A method of measuring eye movemnent using a scieral search coil in a magnetic field. IEEE Transactions on Bio-medical Electronics **10**(4), 137–145 (1963) https://doi.org/10.1109/TBMEL.1963.4322822

Riva, G., Waterworth, J.: Being present in a virtual world, pp. 205–221 (2014)

Segen, V., Avraamides, M.N., Slattery, T.J., Wiener, J.M.: Age-related changes in visual encoding strategy preferences during a spatial memory task. Psychological Research **86**, 404–420 (2022) https://doi.org/10.1007/s00426-021-01495-5

Schrepp, M., Hinderks, A., Thomaschewski, J.: Construction of a benchmark for the user experience questionnaire (ueq). International Journal of Interactive Multimedia and Artificial Intelligence **4**, 40–44 (2017) https://doi.org/10.9781/ijimai.2017.445

Sarupuri, B., Hoermann, S., Whitton, M.C., Lindeman, R.W.: Evaluating and Comparing Game-controller based Virtual Locomotion Techniques. In: ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments (2017). https://doi.org/10.2312/

egve.20171350

Soni, S., Lamontagne, A.: Characterization of speed adaptation while walking on an omnidirectional treadmill. Journal of NeuroEngineering and Rehabilitation **17**(153) (2020) https://doi.org/10.1186/s12984-020-00787-y

Steed, A., Lai, J.: Comparison of hand tracking-based and controller-based interaction in a consumer virtual reality game. Virtual Reality **29**(120) (2025) https://doi.org/10.1007/s10055-025-01190-5

Slater, M.: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society B: Biological Sciences **364**(1535), 3549–3557 (2009) https://doi.org/10.1098/rstb.2009.0138

Starrett, M.J., McAvan, A.S., Huffman, D.J., Stokes, J.D., Kyle, C.T., Smuda, D.N., Kolarik, B.S., Laczko, J., Ekstrom, A.D.: A solution for spatial navigation and memory experiments in virtual reality. Behav Res **53**, 1046–1059 (2021) https://doi.org/10.3758/s13428-020-01481-6

Sacco, K., Ronga, I., Perna, P., Cicerale, A., Del Fante, E., Sarasso, P., Geminiani, G.C.: A virtual navigation training promotes the remapping of space in allocentric coordinates: Evidence from behavioral and neuroimaging data. Frontiers in human neuroscience **16** (2022) https://doi.org/10.3389/fnhum.2022.693968

Serino, S., Stramba-Badiale, C., Magni, G., Castegnaro, A., Riva, G., Tuena, C.: The complementary role of egocentric and allocentric spatial navigation tasks for the diagnosis of alzheimer's disease: A diagnostic meta-analysis. Neuroscience & Biobehavioral Reviews **178**, 106379 (2025) https://doi.org/10.1016/j.neubiorev.2025.106379

Slater, M., Sanchez-Vives, M.V.: Enhancing our lives with immersive virtual reality. Frontiers in Robotics and AI **3** (2016) https://doi.org/10.3389/frobt.2016.00074

Smith, A.J., Tasnim, N., Psaras, Z., Gyamfi, D., Makani, K., Suzuki, W.A., Basso, J.C.: Assessing human spatial navigation in a virtual space and its sensitivity to exercise. JoVE (203) (2024) https://doi.org/10.3791/65332

Theodoropoulos, A., Antoniou, A.: Vr games in cultural heritage: A systematic review of the emerging fields of virtual reality and culture games. Applied Sciences **12**(17) (2022) https://doi.org/10.3390/app12178476

Thornberry, C., Cimadevilla, J.M., Commins, S.: Virtual Morris water maze: Opportunities and challenges. Reviews in the Neurosciences **32**(8), 887–903 (2021) https://doi.org/10.1515/revneuro-2020-0149

Taube, J.S., Valerio, S., Yoder, R.M.: Is navigation in virtual reality with fmri really navigation? Journal of Cognitive Neuroscience **25**(7), 1008–1019 (2013) https://doi.

org/10.1162/jocn_a_00386

Geest, J.N., Frens, M.A.: Recording eye movements with video-oculography and scleral search coils: a direct comparison of two methods. Journal of Neuroscience Methods **114**(2), 185–195 (2002) https://doi.org/10.1016/S0165-0270(01)00527-1

Viaene, P., Vansteenkiste, P., Lenoir, M., Wulf, A.D., Maeyer, P.D.: Examining the validity of the total dwell time of eye fixations to identify landmarks in a building. Journal of Eye Movement Research **9**(3), 1–11 (2016) https://doi.org/10.16910/jemr.9.3.4

Warren, L.E., Bowman, D.A.: User experience with semi-natural locomotion techniques in virtual reality: The case of the virtuix omni. In: Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17), pp. 163–163 (2017). https://doi.org/10.1145/3131277.3134359

Wolbers, T., Hegarty, M.: What determines our navigational abilities? Trends in Cognitive Sciences **14**(3), 138–146 (2010) https://doi.org/10.1016/j.tics.2010.01.001

Wenczel, F., Hepperle, L., Stülpnagel, R.: Gaze behavior during incidental and intentional navigation in an outdoor environment. Spatial Cognition & Computation **17**(1-2), 121–142 (2017) https://doi.org/10.1080/13875868.2016.1226838

Wang, Z., Liu, C., Chen, J., Yao, Y., Fang, D., Shi, Z., Yan, R., Wang, Y., Zhang, K., Wang, H., Wei, H.: Strolling in room-scale vr: Hex-core-mk1 omnidirectional treadmill. IEEE Transactions on Visualization and Computer Graphics **29**(12), 5538–5555 (2023) https://doi.org/10.1109/TVCG.2022.3216211

Xu, T.B., Mostafavi, A., Boot, W.R., Czaja, S., Kalantari, S.: Assessing the feasibility and efficacy of virtual reality navigational training for older adults. Innovation in Aging **9**(1), 099 (2025) https://doi.org/10.1093/geroni/igae099

Yesiltepe, D., Conroy Dalton, R., Ozbil Torun, A.: Landmarks in wayfinding: a review of the existing literature. Cognitive Processing **22**(3), 369–410 (2021) https://doi.org/10.1007/s10339-021-01012-x

Youngstrom, I.A., Strowbridge, B.W.: Visual landmarks facilitate rodent spatial navigation in virtual reality environments. Learning & Memory **19**(3), 84–90 (2012) https://doi.org/10.1101/lm.023523.111

Zhao, H., Frese, L., Venzin, C., Kaszás, D., Weibel, R.P., Hölscher, C., Schinazi, V.R., Thrash, T.: The time course of spatial knowledge acquisition for different digital navigation aids. Computers, Environment and Urban Systems **103**, 101992 (2023) https://doi.org/10.1016/j.compenvurbsys.2023.101992

Zhao, J., Sensibaugh, T., Bodenheimer, B., McNamara, T.P., Nazareth, A., Newcombe, N., Minear, M., Klippel, A.: Desktop versus immersive virtual environments: Effects on spatial learning. Spatial Cognition and Computation **20**(4), 328–363 (2020) https://doi.org/10.1080/13875868.2020.1817925