

Document Version

Final published version

Licence

CC BY

Citation (APA)

Yang, C., Li, D., Dai, Z., & Feng, Y. (2026). Virtual Reality pre-exposure supports wayfinding in a culturally and environmentally unfamiliar multi-level building: short- and long-term evidence from a cross-national study. *Safety Science*, 204, Article 107346. <https://doi.org/10.1016/j.ssci.2026.107346>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Virtual Reality pre-exposure supports wayfinding in a culturally and environmentally unfamiliar multi-level building: short- and long-term evidence from a cross-national study

Chenyi Yang^a, Dewei Li^a, Zhicheng Dai^a, Yan Feng^{b,*}

^a School of Traffic and Transportation, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian District, Beijing 100044, China

^b Delft University of Technology, Department of Transport & Planning, Delft, the Netherlands

ARTICLE INFO

Keywords:

Virtual Reality
Wayfinding
Cross-national
Familiarization
Pre-exposure
Spatial knowledge

ABSTRACT

As public buildings serve increasingly diverse user populations, many pedestrians must navigate environments whose layout, signage, and wayfinding logic are shaped by unfamiliar cultural conventions. This study examined whether Virtual Reality (VR) pre-exposure improves wayfinding in a multi-level Dutch building for participants in China, and whether such benefits persist over time and transfer to new wayfinding tasks. A reference group in the Netherlands completed one VR session, while participants in China completed the same VR wayfinding tasks as baseline, one day later, and again after 14 months. Additional new wayfinding tasks and a supplementary 2D-map pre-exposure condition were introduced to distinguish repeated-route improvement from broader spatial knowledge transfer. We found that, at baseline, participants in China generally required more time, travelled longer distances, and showed higher head-rotation rates than the reference group, suggesting lower efficiency under combined cultural and environmental unfamiliarity. After initial VR exposure, participants in China showed significant reductions in travel time and distance across all original tasks, and much of this improvement remained evident after 14 months. New wayfinding tasks showed that VR familiarization supported route reversal, navigation to new destinations within previously experienced areas, and visual-place recognition, but did not clearly benefit navigation to an unexposed area. Compared with 2D-map pre-exposure, retained VR familiarization produced stronger performance on the original wayfinding tasks and several within-area transfer tasks, but not on the unexposed-area task. These findings suggest that VR pre-exposure can provide durable, route-relevant familiarity, while its transferability remains bounded by the scope of prior virtual exploration.

1. Introduction

With globalization and increasing cross-border mobility, large public facilities such as transportation hubs, hospitals, and university buildings are serving increasingly diverse pedestrian populations. This trend is especially evident in global cities and destinations that attract international students, tourists, and migrant workers. However, the spatial layout, signage systems, and emergency management measures of such facilities are typically designed within a local sociocultural context and are primarily optimized for local users. As a result, a compatibility gap may emerge when these environments are used by people whose spatial expectations, information preferences, and behavioural habits have been shaped elsewhere.

Existing research suggests that cultural background is associated

with differences in wayfinding, including strategy use, risk perception, and wayfinding performance. For example, pedestrians in Japan have been reported to behave more cautiously than those in France and Germany (Sueur et al., 2013; Sprenger et al., 2023), while German pedestrians crossed faster, waited for much shorter periods at the roadside, and accepted slightly larger gaps than Chinese pedestrians (Jiang et al., 2015). Differences in cue use have also been identified between Asian and Western groups (Park 2013). More recent work also found differences in interpersonal distance and pre-crossing movement patterns between Colombian and Portuguese pedestrians (Acela et al., 2025). Australian participants were less willing than Spanish participants to take risks when it was uncertain whether an automated vehicle would yield (Serrano et al., 2024). In wayfinding-related contexts, local conventions can influence side-walking behaviour even when broader route

* Corresponding author.

E-mail address: y.feng@tudelft.nl (Y. Feng).

<https://doi.org/10.1016/j.ssci.2026.107346>

Received 28 March 2026; Received in revised form 29 May 2026; Accepted 28 June 2026

Available online 14 July 2026

0925-7535/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

preferences are shared (Bitgood et al., 2013), and cross-cultural differences in signage preferences further suggest that navigation support cannot always be assumed to function equivalently across user groups (Iftikhar et al., 2021). Under uncertainty, response tendencies may also differ, with some cultural groups showing a stronger tendency to wait for familiar companions before acting (Van der Wal et al., 2024). Cross-cultural differences in walking speed, avoidance patterns, and interpersonal spacing significantly shape pedestrian behaviour and crowd dynamics (Kaminka and Fridman 2018).

Importantly, such differences are not limited to routine navigation but may also affect behaviour in safety-critical situations. Previous studies indicate that cultural background can influence alarm compliance (Jiayu et al., 2025), evacuation route choice (Zhu et al., 2020), and even rescue strategies among firefighters (Chen et al., 2023). For users entering a built environment embedded in an unfamiliar cultural context, the mismatch between spatial expectations formed elsewhere and the design logic of the environment may increase cognitive load and complicate decision-making. In everyday conditions, this may reduce wayfinding efficiency; in emergencies, it may lead to hesitation, route errors, and less efficient use of available egress capacity, thereby increasing operational safety risk.

This challenge is closely related to cultural familiarity with the built environment, which is likely to interact with cultural expectations about building organization and information use. Consistency in culturally shaped design principles, such as axiomaticity, can enhance navigational clarity; however, external visitors are often unfamiliar with these principles (Lu and Bozovic-Stamenovic 2009). Information needs differ between familiar and unfamiliar environments, and accordingly, the ways in which people seek destination-related information also change (Iftikhar and Abbas 2026). When users are unfamiliar with a building, they must rely more heavily on external information than on previously formed mental representations (Raubal 2001). This state is often associated with spatial anxiety (Thomas et al., 2010) and hesitation (Schmitz 1997), both of which can impair wayfinding performance. Unfamiliarity with the environment can also have a negative impact on pedestrians' emotions (Kim et al., 2021), and spatial knowledge is acquired less effectively under pressure or anxiety than in normal conditions (Yu et al., 2023). A study found that participants from mainland China were 2.93 times more likely than local participants to make errors under unfamiliar cultural and traffic conditions (Ye et al., 2020).

Familiarity has a dominant impact on the acquisition of spatial knowledge (Li and Klippel 2016). By contrast, as familiarity increases, pedestrians can form more detailed mental representations of the environment, which improves their ability to identify shorter routes (Muffato and Meneghetti 2020), reduces wrong turns and travel time (Duran 2016), and mitigates the adverse effects of environmental complexity on wayfinding performance (O'Neill 1992). Familiarity also enables a shift from laborious, attention-demanding navigation supported by maps, signs, or instructions to more efficient memory-based movement. For international visitors and other temporary users, however, such familiarity is difficult to acquire through natural exposure before entering the building. Therefore, pre-arrival familiarization delivered through artificial exposure may offer a practical alternative.

Traditional pre-exposure methods, such as studying 2D floor plans, have limited effectiveness, particularly in complex multi-level buildings where vertical connections and three-dimensional spatial relationships are difficult to infer (Meilinger et al., 2006; O'Neill, 1992,). Moreover, compared with familiar users, unfamiliar users are often unable to compensate fully for their lack of spatial knowledge simply by consulting a floor plan (Hölscher et al., 2009). Recently, virtual reality (VR) has shown considerable potential for improving the understanding of spatial structures, especially in domains requiring mental manipulation of complex 3D information (Lin and Suh 2021). Prior research has suggested that 3D forms of representation can improve wayfinding performance relative to more abstract spatial representations such as 2D floor plans (Verghote et al., 2019), and may also support evacuation

success in building environments. Compared with static maps, VR provides dynamic three-dimensional visualization and supports active exploration, both of which are important for spatial learning (Wallet et al., 2009, Münzer and Stahl 2011, Li 2012, Barrett and Hegarty 2016, Sun et al., 2019) and scene recognition (Sauzéon et al., 2016, Afroz et al., 2018). VR can facilitate the acquisition of route and survey knowledge by providing rotational body-based information (He et al., 2019), and high-fidelity virtual environments can support transfer of spatial knowledge to corresponding real-world settings (Wallet et al., 2011). Such an interactive experience may organize spatial knowledge more effectively than passive observation (Al-Alwan and Al-Azzawi 2014).

Based on these advantages, VR can assist in training individuals with navigation difficulties (Riva and Serino 2020). VR can also enhance spatial memory and information retention (Montana et al., 2019, Reggente et al., 2020, Sacco et al., 2022, Toker and Tong 2026). Studies further show that both immersive VR and desktop-based virtual exploration can produce significant transfer of spatial knowledge to subsequent real-world building navigation, although direct real-world experience still tends to yield the best performance (Witmer et al., 1996, Hejtmánek et al., 2020), supporting its use for pre-arrival familiarization. Broader evidence from training research also suggests that immersive technologies can provide a feasible alternative or supplement to conventional training. For example, a recent meta-analysis found that XR-based training can perform comparably to non-simulated training overall (Kaplan et al., 2021). Although this evidence is not specific to wayfinding, it supports the practical rationale for using VR familiarization when direct real-world exposure is difficult, costly, or impractical to provide in advance. Evidence from architectural education shows that immersive VR can effectively support spatial layout learning and retention, even when the virtual models are relatively simple. This suggests that the value of VR familiarization may depend not only on realism, but also on providing structured, distraction-limited opportunities for active spatial exploration (Toker and Tong 2026).

In addition to these learning-related advantages, VR is also a promising tool for safety-related training and behavioural research because it offers a high degree of experimental control and enables precise, automated collection of pedestrian movement and choice data (Feng et al., 2021). In cross-cultural research, VR experiments can also reduce confounding influences arising from differences in climate, infrastructure, and economic conditions, making it possible to conduct equivalent experiments in geographically distant countries. However, the role of VR in cross-cultural wayfinding remains insufficiently understood, and the generalizability of existing findings across cultural contexts is still limited (Hassan et al., 2025).

Despite this growing body of evidence, several questions remain unresolved. First, most VR familiarization studies have examined general environmental unfamiliarity, whereas less is known about situations in which users encounter a building embedded in a different national and cultural context. In such situations, unfamiliarity may arise not only from the absence of prior exposure to the specific building, but also from differences in spatial conventions, signage interpretation, and expectations about building organization. Second, improvements after repeated exposure may reflect memory for specific routes or task sequences rather than transferable spatial learning. Third, although VR is often proposed as a pre-arrival familiarization tool, its added value relative to simpler forms of spatial pre-exposure, such as studying a 2D floor plan, remains insufficiently examined. Finally, evidence on the long-term retention of VR-based wayfinding benefits remains limited, especially over intervals longer than a few days or weeks.

To address these gaps, the present study builds on a previous VR wayfinding experiment conducted in the Netherlands using a virtual model of a Dutch university building (Feng et al., 2022a). Using the same virtual environment, software, procedure, and equipment, we conducted corresponding experiments with participants in China who were unfamiliar with both the building and its local spatial context. The

Netherlands reference group (NL) completed one session and served as a local reference group with mixed levels of prior building familiarity, whereas participants in China completed an initial VR session (T1), a repeated session on the following day (T2), and a retention session approximately 14 months later (T3). To examine whether the observed improvements extended beyond repeated-route learning, additional new wayfinding tasks were administered after T3. To provide a non-immersive comparator, an additional group of participants received 2D-map pre-exposure before completing either the original or the new wayfinding tasks. This design allows us to examine: (1) baseline cross-national differences under combined cultural and environmental unfamiliarity; (2) short-term and long-term changes following VR exposure; (3) the extent to which learning transfers to new tasks within or beyond the previously experienced area; and (4) how retained VR familiarization compares with 2D-map pre-exposure. The study therefore does not attempt to isolate culture as a single causal factor, but instead evaluates VR-based familiarization as a practical intervention for users encountering an unfamiliar multi-level building in a different national and cultural context.

The remainder of this paper is organized as follows. Section 2 describes the experimental design, participants, tasks, measures, and statistical procedures. Section 3 presents the results, including the baseline cross-national comparison, short- and long-term VR familiarization effects, new wayfinding transfer tasks, and comparisons with the 2D-map pre-exposure condition. Sections 4 and 5 discuss the findings and conclude the paper.

2. Materials and methods

2.1. Study overview and experimental design

Previous studies have shown that wayfinding performance in VR is broadly comparable to that in real environments (Dong et al., 2022) and significantly correlated with it, with stronger associations under more difficult virtual tasks (Coutrot et al., 2019). A recent validation study in a multi-level university building reported that pedestrian wayfinding in an immersive VR replica reproduced the main spatial strategies and overall time-efficiency patterns observed in the real building (van Beek et al., 2024). Therefore, we designed a VR experiment to simulate a culturally unfamiliar environment based on a real multi-level building. This study consisted of four related components: a local-reference VR session conducted in the Netherlands, a three-wave longitudinal VR experiment conducted in China, a set of new wayfinding tasks administered to the VR participants in China, and an additional 2D-map pre-exposure comparator.

The virtual environment represented the Faculty of Civil Engineering and Geosciences building at Delft University of Technology in the Netherlands. The same virtual model, WayR software (Feng et al., 2022b), task instructions, and VR equipment were used in the Netherlands and China. The design followed four analytical steps. First, baseline cross-national differences were examined by comparing the participants in China at T1 with the Netherlands reference group (NL). Because the two groups differed not only in national/cultural background but also in prior environmental familiarity, this comparison was interpreted as a contrast under combined cultural and environmental unfamiliarity rather than as an isolated test of cultural effects. Second, short-term familiarization effects were examined by comparing T1 and T2 within the group in China. Third, long-term retention was assessed by comparing T3 with T1 and T2. Fourth, additional new wayfinding tasks and a 2D-map comparator were introduced to examine whether the observed benefits extended beyond repeated-route learning and whether immersive VR familiarization differed from non-immersive floor-plan pre-exposure.

This design supports descriptive and exploratory comparisons of short-term and long-term changes following VR exposure, and provides an exploratory benchmark against 2D-map familiarization. However, it

does not isolate culture as an independent causal factor, nor does it provide a pure no-exposure control condition for repeated testing effects.

2.2. Virtual environment and original wayfinding tasks

The virtual environment was developed using WayR, a VR platform designed for studying pedestrian wayfinding and evacuation behaviour in multi-level buildings (Feng et al., 2022b). The model represented a four-floor Dutch university building. Each floor contained two parallel corridors, several intersections, four staircases, and four elevators. The ground floor contained five main exits. Rooms with odd and even numbers were distributed along different corridors, and participants could switch between corridors at designated decision points. A path segment was defined as the corridor section between two adjacent decision points along the two main parallel corridors, as shown in Fig. 1 and Fig. 2.

Each VR session included four original wayfinding tasks. Task 1 was a same-floor wayfinding task from Room 4.02 to Room 4.99. Tasks 2 and 3 were between-floor wayfinding tasks from Room 4.99 to Room 2.01 and from Room 2.01 to Room 4.64, respectively. Task 4 was an evacuation-oriented task. In this task, an alarm and an on-screen instruction were presented, and participants were instructed to leave the building as quickly as possible. Because the task did not include smoke, crowding, physical threat, or social interaction, it should be interpreted as an emergency-oriented wayfinding scenario rather than a full evacuation simulation.

2.3. Participants and experimental conditions

A total of 36 participants were included in the Netherlands reference group (NL), and 34 participants were included in the longitudinal VR group in China. The participants in China completed T1, T2, and T3. All participants in China were students and reported no prior familiarity with the virtual building. The participants in the Netherlands had mixed levels of prior familiarity with the faculty building, ranging from not at all familiar to very familiar. This group was not intended to represent a uniformly building-familiar control group; rather, it reflected an ecologically realistic local-user population with mixed levels of prior familiarity with the building.

Table 1 presents the demographic characteristics of the two groups and the corresponding statistical test results. The participants in China were aged 22–28 years ($M = 23.44$, $SD = 1.31$), whereas the participants in the Netherlands were aged 19–66 years ($M = 28.56$, $SD = 10.30$). The age difference between the two groups was statistically significant. The age comparison between the two participant groups is shown in Fig. 3. To examine whether this age difference influenced the baseline cross-national comparison, an age-restricted sensitivity analysis was conducted by excluding Netherlands participants older than 30 years (subsample NL1, 29 participants). In this subsample, the age difference between groups was no longer statistically significant. A second sensitivity analysis was conducted by excluding participants in the Netherlands who reported medium or higher familiarity with the building (subsample NL2, 25 participants). These sensitivity analyses were used to evaluate the robustness of the baseline findings, but they do not fully remove all demographic and familiarity-related differences between groups. However, educational background differed between participants in China and the Netherlands and could not be fully controlled because of the limited sample size and sparse distributions across education categories.

In addition to the VR groups, 40 new Chinese participants were recruited for the 2D-map pre-exposure comparison. They were randomly assigned to two subgroups: one subgroup completed the original tasks after map pre-exposure (2DO group), and the other subgroup completed the new wayfinding tasks after map pre-exposure (2DN group). Demographic comparisons indicated no significant differences between the

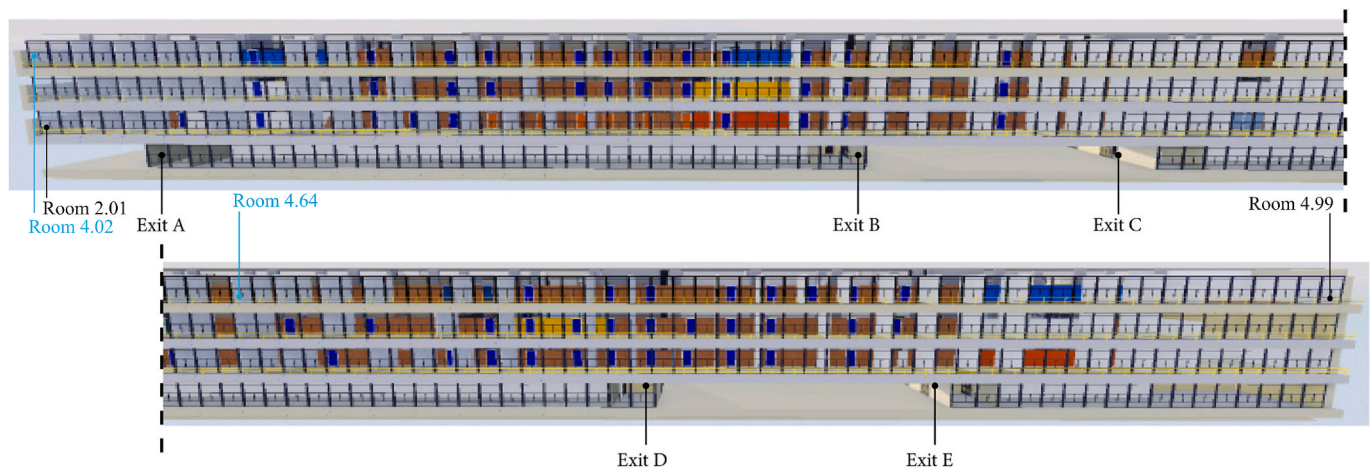


Fig. 1. Overview of the virtual building (blue room numbers indicate rooms located in the covered corridor on the opposite side). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

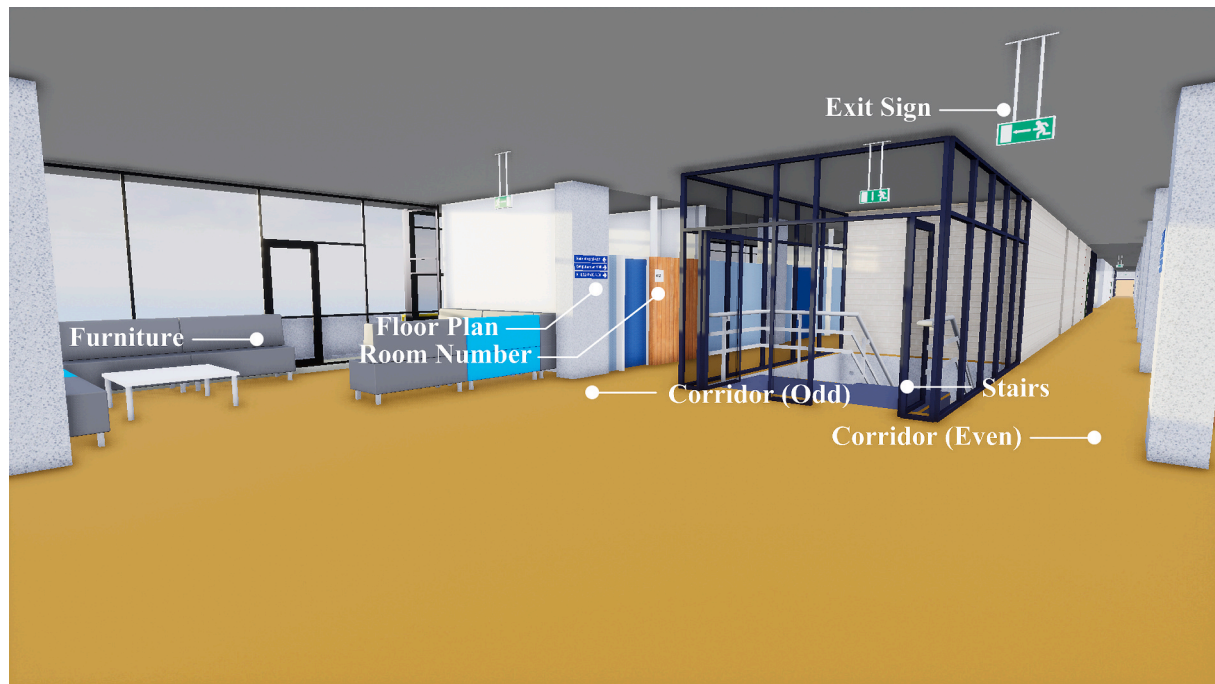


Fig. 2. Indoor virtual environment.

2D-map participants and the VR participants in China, nor between the two 2D-map subgroups (see Table A. 6 and Table A. 7).

2.4. Experimental procedure

For each VR session, participants first read the experimental instructions and signed an informed consent form. They then wore the head-mounted display and entered a virtual tutorial room to become familiar with the basic movement and interaction controls. After the tutorial, participants completed the four original wayfinding tasks consecutively. No breaks were provided between tasks. After completing the session, participants filled out user-experience questionnaires.

The Netherlands reference group completed one VR session. The VR group in China completed the same task sequence at three time points. T1 was the initial exposure session. T2 was conducted on the following day and was used to assess short-term change after prior VR exposure. T3 was conducted approximately 14 months later and was used to assess

long-term retention. During the interval between T2 and T3, the research team provided no additional exposure to the virtual environment. Participants were also asked whether they had revisited or otherwise familiarized themselves with the experimental scene during the interval.

Two months after T3, the VR participants in China were invited to complete an additional new wayfinding session (NT). Due to scheduling constraints, 33 of the 34 participants completed this session.

2.5. New wayfinding tasks

The original T1, T2, and T3 sessions used the same four tasks in the same order. Therefore, performance improvement across sessions could partly reflect memory for specific routes, destinations, or task order. To examine whether the learning effect extended beyond repeated-route performance, four new wayfinding tasks (NT) were designed for the VR participants in China and administered two months after T3.

Table 1
Demographic characteristics of the participants in NL and China (T1/T2/T3).

Descriptive information	Category	The Netherlands	China	Test result
Gender	Male	17, 47.2%	18, 52.9%	$p = 0.632$ $\chi^2(1, N = 70) = 0.23, \phi = 0.06$ $p = 0.002$ Cramér's $V = 0.444$
	Female	19, 52.8%	16, 47.1%	
Highest education level	High school or equivalent	5, 13.9%	0, 0.0%	$p = 0.226$ Cramér's $V = 0.257$
	Bachelor's degree or equivalent	16, 44.4%	29, 85.3%	
	Master's degree or equivalent	11, 30.6%	4, 11.8%	
	Doctoral degree or equivalent	4, 11.1%	1, 2.9%	
Previous experience with VR	Never	13, 36.1%	6, 17.6%	$p = 0.662$ Cramér's $V = 0.185$
	Seldom	14, 38.9%	18, 52.9%	
	Sometimes	6, 16.7%	9, 26.5%	
	Often	3, 8.3%	1, 2.9%	
Familiarity with any computer gaming	Not at all familiar	4, 11.0%	5, 14.7%	$p = 0.222$ Cramér's $V = 0.242$
	little familiar	11, 30.6%	7, 20.5%	
	Moderately familiar	5, 13.9%	9, 26.5%	
	Quite-a-bit familiar	11, 30.6%	9, 26.5%	
	Very familiar	5, 13.9%	4, 11.8%	
Frequency of Emergency drill	Never	0, 0.0%	1, 2.9%	$p < 0.001$ Cramér's $V = 0.770$
	Seldom	9, 25.0%	10, 29.4%	
	Sometimes	18, 50.0%	20, 58.9%	
	Often	9, 25.0%	3, 8.8%	
Familiarity with the faculty building	Not at all familiar	9, 25.0%	34, 100.0%	
	little familiar	16, 44.4%	0, 0.0%	
	Moderately familiar	5, 13.9%	0, 0.0%	
	Quite-a-bit familiar	4, 11.1%	0, 0.0%	
	Very familiar	2, 5.6%	0, 0.0%	

The four new tasks probed different levels of transfer. New Task 1 was a reverse-route task derived from original Task 3 and required participants to navigate from Room 4.64 to Room 2.01. New Task 2 required participants to navigate to a new destination within the previously experienced area, from Room 2.01 to Room 3.77.1. New Task 3 required participants to navigate from Room 3.76 to Exit B, a destination outside the experienced area. New Task 4 was a landmark-based task. Participants were shown a photograph of a location inside the building and were asked to navigate from Exit B to that location.

A special feature of New Tasks 3 and 4 was the use of Exit B, which had not been encountered during the original VR sessions. The building contained three exit groups, namely A/B, C/D, and E, all located on the first floor. These exit groups were mutually invisible and were not directly connected on the first floor; moving from one exit group to another required first going upstairs, then moving horizontally, and finally descending again. In the original evacuation task, participants selected exits C or D, and each trial ended immediately upon arrival at the chosen exit. Therefore, participants had not previously reached or seen Exit B during the original experimental sessions.

The same behavioural measures as in the original tasks were recorded for the new wayfinding tasks, including travel time, travel distance,

and head rotation. Because the new wayfinding tasks differed from the original tasks in route length and spatial complexity, performance was also expressed using a Best Performance Ratio (BPR).

2.6. 2D-map pre-exposure comparator

To compare immersive VR familiarization with a non-immersive form of spatial pre-exposure, a 2D-map condition was introduced as a comparator. Forty new Chinese participants were recruited and randomly assigned to two subgroups. The 2DO subgroup (20 participants) completed the original four tasks after map pre-exposure, whereas the 2DN subgroup (20 participants) completed the four new wayfinding tasks after map pre-exposure.

The 2D-map material consisted of top-down floor plans of the four floors of the same building. The map included the room numbers used in the original tasks, the labels of Exits, and textual labels indicating staircases. Elevators were not shown because elevator use was not allowed. The map therefore provided global layout information but did not provide a first-person visual experience of corridors, intersections, or landmarks.

The duration of map pre-exposure was 668 s, matching the average duration of the initial VR familiarization session in T1. During map pre-exposure, participants were first instructed to mark the routes of the four original tasks on the floor plan in sequence. They then used the remaining time to freely inspect and memorize the map. One week after map pre-exposure, participants completed either the original tasks or the new wayfinding tasks in the same VR environment.

The 2D-map comparator served two purposes. First, the 2DO subgroup allowed comparison between map pre-exposure and retained VR familiarization on the original tasks. Second, the 2DN subgroup allowed comparison between map pre-exposure and VR familiarization on the new wayfinding tasks. Because the 2D-map condition involved a single pre-exposure session followed by testing after one week, whereas the VR group had completed repeated immersive sessions over a longer period, these comparisons should be interpreted as pragmatic comparisons between familiarization formats rather than as a fully controlled test of the isolated effect of immersion.

2.7. Data collection and measures

WayR recorded timestamped data at 10 Hz, including three-

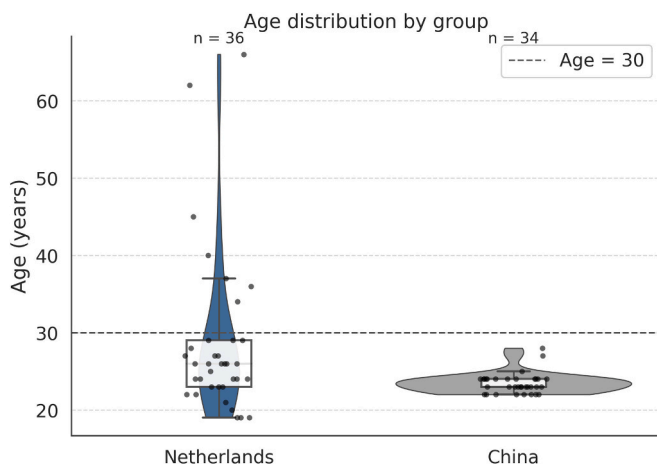


Fig. 3. Age distribution by group.

dimensional position coordinates and head orientation. These data were used to derive three categories of measures: wayfinding decision-making, head rotation, and wayfinding performance.

Wayfinding decision-making included strategy choice, staircase choice, exit choice, path usage, and decision-point usage. Three wayfinding strategies were distinguished: the central-point strategy, the directional strategy, and the floor strategy (Hölscher et al., 2006). In Task 1, the directional strategy referred to first moving toward the corridor containing the destination room, whereas the central-point strategy referred to first moving toward the corridor containing the starting room. In Tasks 2 and 3, the directional strategy referred to prioritizing horizontal movement before changing floors, whereas the floor strategy referred to prioritizing vertical movement. Strategy was not classified for Task 4 because this task focused on staircase and exit choice under an emergency-oriented instruction.

Path usage and decision-point usage were summarized descriptively. Because participants could traverse the same path or decision point multiple times, these counts were not independent observations at the participant level. They were therefore used to provide spatial context for the route-choice and performance results rather than as independent inferential outcomes.

Head rotation was calculated as the average angular change in yaw orientation (left–right head-turning activity) per second during task execution (Zhang et al., 2021). This measure may reflect visual scanning, route checking, turning behaviour, detours, or disorientation, but it does not directly identify what participants looked at or reveal their cognitive state. The average rotation change $\bar{Y}$ per task was derived from Formulas (1) and (2), with $Y(t)$ as the instantaneous rotation change, $Y(t)$ the yaw coordinate at time t , dt the time step, and T the task duration (Feng et al., 2022b).

$$Y(t) = \min(360 - |Y_{t+dt} - Y_t|, Y_{t+dt} - Y_t)$$

$$\bar{Y} = \frac{\sum_{t=1}^T Y}{T}$$

Wayfinding performance was assessed using travel time and travel distance. Travel time (T) is defined as the duration from start to destination for each task. Travel distance, calculated from 3D coordinates using Formulas (3) and (4) with a sampling interval of $dt = 0.1$ s, reflects the actual path length. Both corridor (D_c) and staircase (D_s) travel distance are computed (Feng et al., 2022b).

$$D_c = \sum_{t=1}^{T_c} \sqrt{(x_{t+dt} - x_t)^2 + (y_{t+dt} - y_t)^2}$$

$$D_s = \sum_{t=1}^{T_s} \sqrt{(x_{t+dt} - x_t)^2 + (y_{t+dt} - y_t)^2 + (z_{t+dt} - z_t)^2}$$

For new wayfinding tasks and cross-task comparisons involving different route lengths or spatial complexity, performance was additionally expressed as a Best Performance Ratio (BPR). BPR was calculated as each participant's score divided by the mean score of three experimenters who were fully familiar with the building and completed the same task under the same experimental conditions. Values closer to 1 indicate performance closer to the expert-reference level. For travel time and travel distance, lower values indicate more efficient performance. For head rotation, the ratio should be interpreted as relative head-turning activity rather than as a direct indicator of optimality.

User experience was assessed using face validity, simulator sickness, presence, and usability measures. Face validity was rated on a 5-point Likert scale (Jamieson 2004). Simulator sickness was measured using the Simulator Sickness Questionnaire (Kennedy et al., 1993). Presence was measured using the Presence Questionnaire (Witmer et al., 2005), and usability was assessed using the System Usability Scale (Brooke 1996).

2.8. Statistical analysis

Categorical outcomes, including wayfinding strategy, staircase choice, and exit choice, were analysed using Pearson's chi-square tests or Fisher's exact tests for between-group comparisons, depending on expected cell counts. For paired categorical comparisons between T1 and T2, McNemar's tests were used for binary variables and Stuart-Maxwell tests for variables with more than two categories. For three-wave comparisons of binary strategy variables across T1, T2, and T3, Cochran's Q tests were used. Staircase and exit choices across three sessions were summarized descriptively when the number of categories and sparse cell counts made repeated nominal tests unsuitable.

Continuous outcomes, including travel time, travel distance, and head rotation, were analysed according to distributional assumptions and repeated-measures structure. Normality was assessed using the Shapiro-Wilk test. Independent-samples t-tests or Mann-Whitney U tests were used for between-group comparisons. Paired-samples t-tests or Wilcoxon signed-rank tests were used for two-wave within-subject comparisons. For three-wave within-subject comparisons, Friedman tests were used as omnibus tests, followed by post hoc pairwise comparisons where appropriate.

Holm-adjusted p-values were used to control the family-wise error rate within each family of related comparisons. A family was defined as the set of task-specific comparisons addressing the same inferential contrast for the same outcome variable. For example, the four task-specific comparisons of travel time for T1 versus T2 were treated as one family, whereas the four task-specific comparisons of travel time for T1 versus the Netherlands reference group were treated as a separate family. This approach controlled for multiple task-level tests while avoiding the combination of conceptually distinct outcomes and research questions into a single overly conservative correction.

Effect sizes were reported as ϕ or Cramér's V for categorical tests, Cohen's d/dz for t-tests, r for Mann-Whitney U and Wilcoxon signed-rank tests, and Kendall's W for Friedman tests. Because the primary interest was in effect magnitude rather than direction, absolute values are reported for signed effect-size measures (i.e. $|dz|$ and $|r|$). Statistical significance was set at $p < 0.05$.

A sensitivity/sample-size assessment was conducted using G*Power 3.1. The available sample size was adequate for detecting large between-group effects but not for reliably detecting small-to-medium effects, especially in subgroup analyses and task-specific comparisons. Therefore, non-significant results should not be interpreted as evidence of no difference or equivalence.

3. Results

3.1. Baseline cross-national comparison (China T1 vs the Netherlands)

3.1.1. Wayfinding decision-making

Strategy distributions across tasks are summarized in Table 2. In Task 1, the Netherlands reference group was more likely to adopt a directional strategy than participants in China at T1. This difference was not significant in either the NL1 or NL2 group. However, the relatively small p-value observed for the NL1 group suggests that the potential influence of age can be ruled out to a limited extent. The statistical results are presented in Table 3.

Fig. 4, Fig. 5, Fig. 6, and Fig. 7 present aggregated trajectory patterns for participants in China at T1 and the Netherlands reference group. These trajectory counts were not subjected to inferential testing because multiple path-segment observations within a task were not independent at the participant level. Therefore, the trajectory figures are intended to provide descriptive spatial context for the route-choice and performance results rather than independent statistical evidence. Overall route structures were broadly similar across groups, although the intensity of use varied across specific corridors and decision points. In particular, Decision Point 4 (the second from the right in Fig. 4) was traversed much

Table 2

Wayfinding strategy distributions across groups and sessions.

Nation	Group	Task 1		Task 2		Task 3	
		Central-point	Directional	Directional	Floor	Directional	Floor
The Netherlands	NL	14 (38.9%)	22 (61.1%)	3 (8.3%)	33 (91.7%)	2 (5.6%)	34 (94.4%)
	NL1 (under 30)	12 (41.4%)	17 (58.9%)	2 (6.90%)	27 (93.1%)	2 (6.9%)	27 (93.1%)
	NL2 (less familiar)	11 (44.0%)	14 (56.0%)	3 (12.0%)	22 (88.0%)	0 (0%)	25 (100%)
China	T1 (Day 0)	22 (64.7%)	12 (35.3%)	12 (35.3%)	22 (64.7%)	1 (2.9%)	33 (97.1%)
	T2 (Day 1)	23 (67.6%)	11 (32.4%)	5 (14.7%)	29 (85.3%)	1 (2.9%)	33 (97.1%)
	T3 (after 14 months)	22 (64.7%)	12 (35.3%)	8 (23.5%)	26 (76.5%)	0 (0%)	34 (100%)

Table 3

Test result of wayfinding strategy (T1 vs NL/NL1/NL2).

Group	Task 1	Task 2	Task 3
	Test result (Chi-square)	Test result (Chi-square)	Test result (Fisher exact)
NL vs. T1	$\chi^2(1, N = 70) = 4.67, p = 0.031, \phi = 0.26$	$\chi^2(1, N = 70) = 7.55, p = 0.006, \phi = 0.33$	$p = 1.000$, Cramér's V = 0.07
NL1 vs. T1	$\chi^2(1, N = 63) = 3.43, p = 0.064, \phi = 0.23$	$\chi^2(1, N = 63) = 7.30, p = 0.007, \phi = 0.34$	$p = 0.590$, Cramér's V = 0.09
NL2 vs. T1	$\chi^2(1, N = 59) = 2.51, p = 0.113, \phi = 0.21$	$\chi^2(1, N = 59) = 4.12, p = 0.042, \phi = 0.26$	$p = 1.000$, Cramér's V = 0.11

less frequently than the other decision points.

In Task 2, participants in the Netherlands reference group were more likely to adopt a floor strategy than participants in China at T1; 91.7% of participants in the Netherlands reference group chose to descend first. The same difference was also observed in both the NL1 and NL2 groups, where it reached statistical significance. Staircase choice (see in Table 4) also differed significantly between NL and T1, with staircase use being more dispersed among participants in China. No significant differences were found between T1 and NL1 or between T1 and NL2, although the comparison with NL1 showed a marginal tendency, $p = 0.084$. This finding helps reduce, although not eliminate, the possibility that the observed difference was driven by age. In addition, the significant difference observed in the full NL group suggests that the effect may reflect the combined influence of cultural background and environmental familiarity. In Tasks 3 and 4, no significant between-group differences were found in wayfinding strategy, staircase choice, or exit choice when T1 was compared with NL, NL1, or NL2.

3.1.2. Wayfinding performance and head rotation

Table 5 summarizes wayfinding performance (travel time and travel distance) and head rotation at baseline. Across tasks, participants in China at T1 generally exhibited longer travel times and greater travel

distances than the Netherlands reference group, indicating lower wayfinding efficiency during initial exposure to the virtual building. Participants in China also tended to show higher head-rotation rates, although statistically significant between-group differences in head rotation were not observed in every task. These differences should be interpreted as differences in head-turning activity rather than as direct evidence of differences in visual attention or cognitive uncertainty. Separate comparisons were conducted between T1 and NL1 and between T1 and NL2. The indicators showing significant between-group differences remained unchanged. This suggests that age was unlikely to be the primary driver of the observed effects. However, because participants who were “a little familiar” with the environment were still included, the potential influence of environmental familiarity cannot be fully ruled out. The detailed statistical results are provided in Table A. 1 and Table A. 2.

3.1.3. User experience

Face validity, presence, usability, and simulator sickness were examined to assess user experience in the VR environment. No statistically significant between-group differences were observed in the overall scores reported in Table A. 3, indicating broadly comparable overall acceptability of the VR procedure for participants in China and the

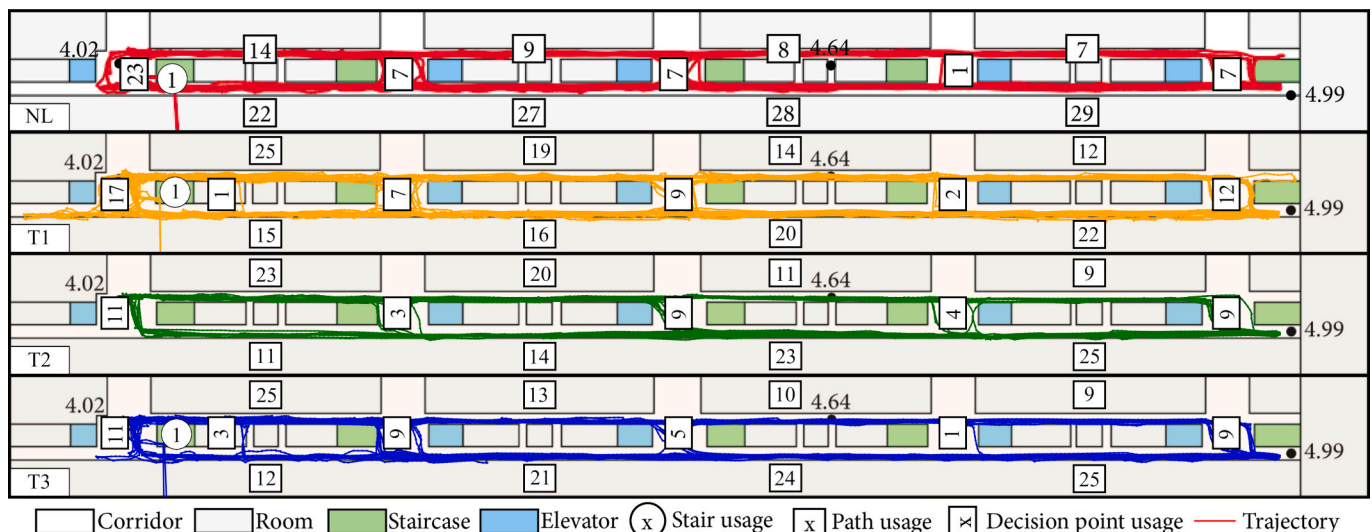


Fig. 4. Aggregated trajectories in Task 1. Colors indicate different participant groups, as labelled on the left side of each subfigure (same below).

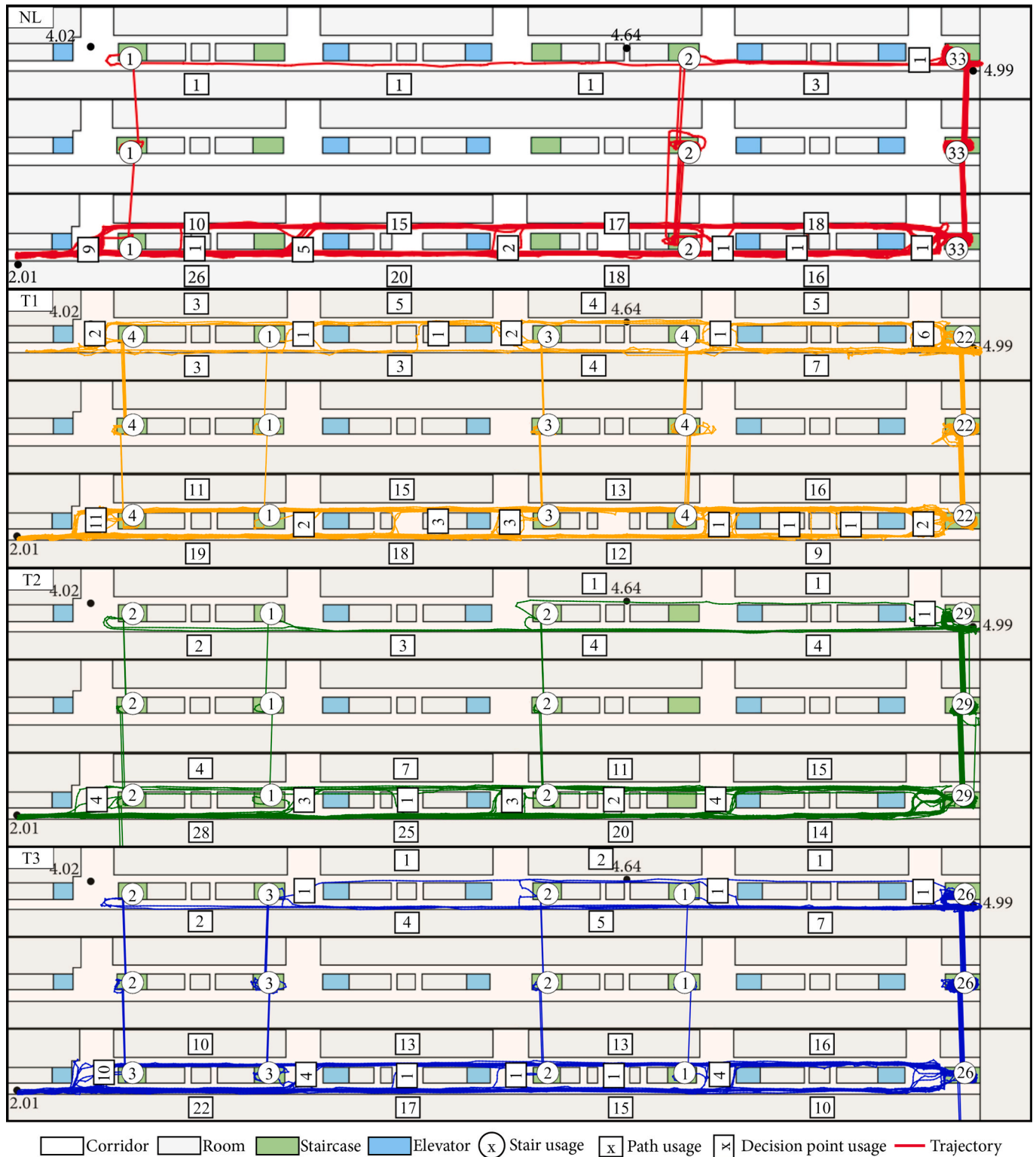


Fig. 5. Aggregated trajectories in task 2.

Netherlands. Several subscale- and item-level differences were observed; however, after Holm correction, only the difference in interface quality remained significant.

3.1.4. Summary

Taken together, the baseline comparison indicates both commonalities and cross-group differences in VR wayfinding in an unfamiliar national and environmental context. Participants in the Netherlands

reference group were more likely to rely on directional strategies in Task 1 and floor strategies in Task 2, whereas participants in China at T1 tended to travel longer distances, require more time, and show higher head-rotation rates, consistent with lower wayfinding efficiency and more intensive visual scanning during first exposure. For user experience, no significant between-group differences were observed at the overall level.

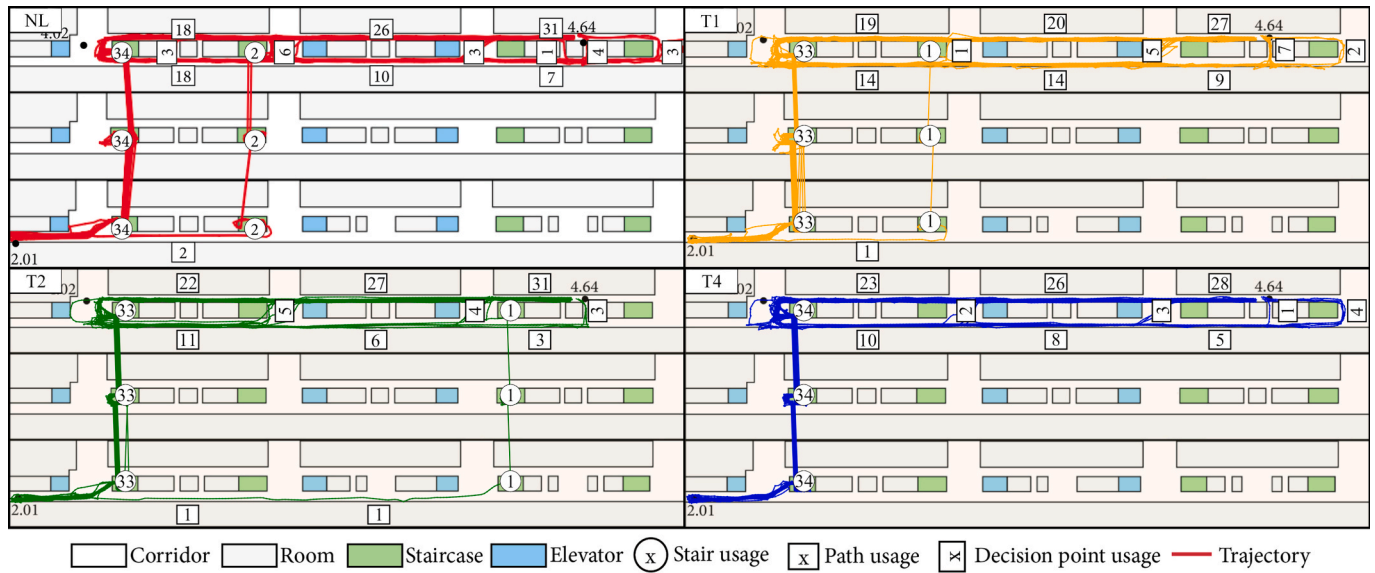


Fig. 6. Aggregated trajectories in task 3.

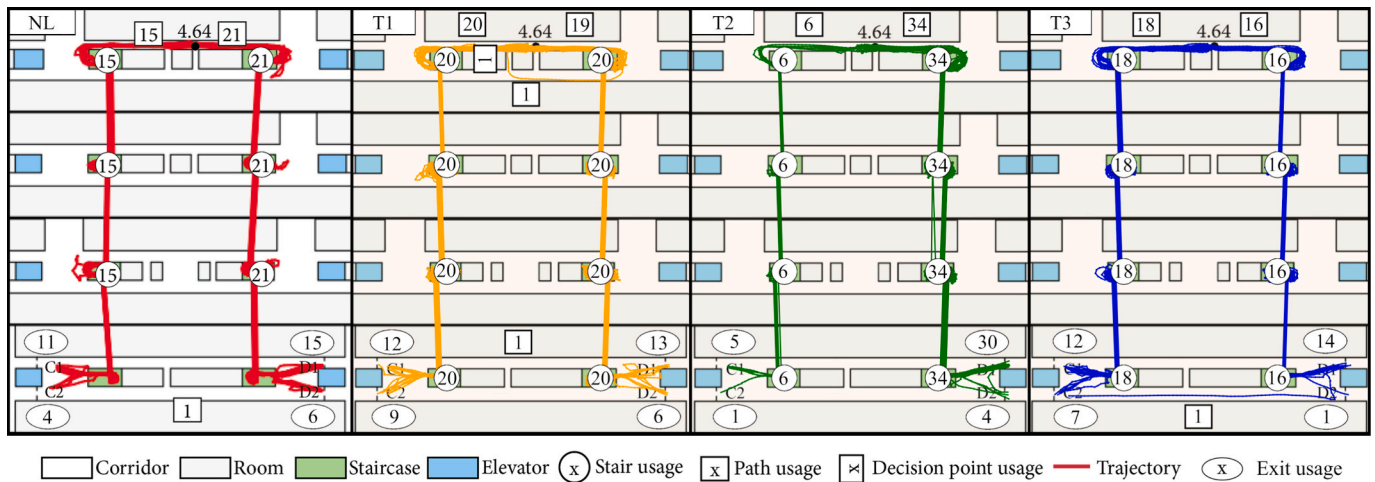


Fig. 7. Aggregated trajectories in task 4.

Table 4

Test result of staircase and exit choice (T1 vs NL/NL1/NL2).

Group	Task 2	Task 3	Task 4
	Staircase choice (Fisher exact test)	Staircase choice (Fisher exact test)	Staircase choice (Chi-square test)
NL vs. T1	$p = 0.050$, Cramér's $V = 0.35$	$p = 1.000$, Cramér's $V = 0.07$	$\chi^2(1, N = 70) = 1.42$, $p = 0.234$, $\phi = 0.14$
NL1 vs. T1	$p = 0.084$, Cramér's $V = 0.35$	$p = 0.590$, Cramér's $V = 0.09$	$\chi^2(1, N = 63) = 1.32$, $p = 0.251$, $\phi = 0.15$
NL2 vs. T1	$p = 0.197$, Cramér's $V = 0.33$	$p = 0.569$, Cramér's $V = 0.11$	$\chi^2(1, N = 59) = 0.81$, $p = 0.367$, $\phi = 0.15$
			Exit choice (Chi-square test)
NL vs. T1			$\chi^2(3, N = 70) = 3.42$, $p = 0.332$, Cramér's $V = 0.22$
NL1 vs. T1			$\chi^2(3, N = 63) = 3.20$, $p = 0.362$, Cramér's $V = 0.23$
NL2 vs. T1			$\chi^2(3, N = 59) = 3.67$, $p = 0.299$, Cramér's $V = 0.25$

3.2. Short-term familiarization effects among participants in China (T1-T2)

This section examines whether a single prior exposure to the virtual environment improved wayfinding and evacuation performance among participants in China. Within-subject comparisons between T1 and T2 were used to assess short-term familiarization effects.

3.2.1. Wayfinding decision-making

Strategy use did not differ between T1 and T2 in Task 1 (McNemar's test, exact $p = 1.000$; continuity-corrected $\chi^2(1) = 0.00$, $p = 1.000$),

Task 2 (McNemar's test, exact $p = 0.119$; continuity-corrected $\chi^2(1) = 2.40$, $p = 0.121$) or Task 3 (McNemar's test, exact $p = 1.000$; continuity-corrected $\chi^2(1) = 0.50$, $p = 0.480$). The number of participants adopting each strategy are presented in Table 2.

Table 6 presents the results of the Stuart-Maxwell tests for staircase choice and exit choice. No significant change in staircase choice was found from T1 to T2 in Task 2 or Task 3. However, significant within-subject changes were observed in the evacuation task for both staircase choice and exit choice. At T2, participants more often selected the staircase located ahead of their initial position rather than the one behind them.

Table 5
Means and standard deviations of wayfinding performance and head rotation for participants in China at T1 and The Netherlands reference group.

Item	Task	The Netherlands Mean(SD)	China Mean (SD)	Test	p-value	p-holm	t or U	effect size
Travel time (s)	1	172.40 (21.85)	198.17 (44.48)	U	0.010	0.030	830.00	0.31
	2	209.45 (26.86)*	243.61 (44.02)	U	<0.001	<0.001	1007.00	0.55
	3	144.25 (22.60)	154.30 (30.54)	U	0.062	0.124	771.00	0.22
	4	68.35(13.52)	72.21 (13.53)	U	0.136	0.136	739.00	0.18
Travel distance (m)	1	194.22 (10.26)	208.80 (33.73)	U	0.100	0.100	752.00	0.20
	2	238.62 (8.35)	262.30 (36.96)	U	<0.001	<0.001	1078.00	0.65
	3	166.79 (16.20)	169.94 (12.17)	U	0.014	0.028	822.00	0.29
	4	72.45 (3.79)*	75.75 (4.82)	U	0.003	0.009	866.00	0.36
Head rotation (°/s)	1	10.60 (4.52)	12.62 (5.54)*	U	0.136	0.159	739.00	0.18
	2	12.98 (2.92)	15.76 (3.69)*	U	<0.001	<0.001	912.00	0.42
	3	14.34 (2.82)*	15.85 (3.94)*	T	0.074	0.159	−1.82	−0.44
	4	27.34 (5.48)*	30.30 (6.87)*	T	0.050	0.159	−1.97	−0.47

* indicates normal distribution. In the column of p-value, ‘T’ represents Independent Samples or Paired T-Test, and ‘U’ represents Mann-Whitney U Test (same below).

Table 6
Stuart-Maxwell test results of staircase and exit choice (T1-T2).

Task	Item	χ^2	asymptotic p	permutation p	random swaps
2	Staircase choice	$\chi^2(4) = 5.20$	0.267	0.290	50,000
3	Staircase choice	$\chi^2(4) = 2.00$	0.736	1.000	
4	Staircase choice	$\chi^2(4) = 11.27$	0.024	0.001	
4	Exit choice	$\chi^2(3) = 15.45$	0.001	<0.001	

3.2.2. Wayfinding performance and head rotation

Changes in travel time, travel distance, and head rotation from T1 to T2 are summarized in Table 7. Across all four tasks, participants in China showed significantly lower travel time and shorter travel distance at T2 than T1, indicating improved wayfinding efficiency after one prior exposure to the virtual environment. For head rotation, a significant reduction was observed in Task 1 and Task 2, whereas no significant change was found in Task 3 or Task 4. The reduction in Tasks 1 and 2 suggests less head-turning activity after short-term familiarization. This may be consistent with reduced visual scanning, fewer route-checking behaviours, or more direct movement, but it should not be interpreted as a direct measure of cognitive uncertainty.

3.2.3. Summary

Overall, the T1-T2 comparison indicates a clear short-term familiarization effect among participants in China, reflected in improved navigation efficiency across all tasks and reduced head rotation in the earlier wayfinding tasks. In addition, significant within-subject changes were observed in staircase and exit choice in Task 4.

3.3. Long-term retention among participants in China (T1-T2-T3)

To evaluate the durability of VR-based familiarization, participants in China completed a follow-up session approximately 14 months after T1 and T2. The T1-T2-T3 comparison was used to assess whether initial gains were retained over time and whether any performance decline occurred relative to the short-term retest.

3.3.1. Wayfinding decision-making

For binary wayfinding-strategy choices, Cochran’s Q tests indicated no significant change across T1, T2, and T3 for Task 1 ($Q = 0.095$, $p = 0.953$), Task 2 ($Q = 3.895$, $p = 0.143$), and Task 3 ($Q = 1.000$, $p = 0.607$). This suggests that participants in China adopted broadly stable wayfinding strategies across the three sessions (see Table 2 for detailed counts).

Cochran’s Q test is designed for repeated dichotomous outcomes and was therefore not applicable to staircase choice and exit choice, which involved more than two nominal response categories. In addition, some categories contained small cell counts. Therefore, staircase and exit choices are summarized descriptively in Fig. 8. The Sankey diagrams should be interpreted as visualizations of observed choice-transition patterns rather than as inferential tests of statistically significant

Table 7
Short-term within-subject comparisons of travel time, travel distance, and head rotation for participants in China (T1 vs. T2).

Item	Task	T1 Mean(SD)	T2 Mean(SD)	Test	p-value	p-holm	t or W	Cohen’s dz or r
Travel time (s)	1	198.17(44.48)	155.44(15.89)	W	<0.001	<0.001	21.00	0.811
	2	243.61(44.68)	204.64(34.44)	W	<0.001	<0.001	20.00	0.814
	3*	154.30(31.00)	142.29(28.15)	T	<0.001	<0.001	−3.937	−0.675
	4*	72.21(13.73)	65.81(18.15)	T	0.002	0.006	−3.376	−0.579
Travel distance (m)	1	208.80(34.23)	185.77(1.60)	W	<0.001	<0.001	24.00	0.802
	2	262.30(37.51)	238.17(6.68)	W	<0.001	<0.001	19.00	0.817
	3	169.94(12.36)	161.80(5.39)	W	<0.001	<0.001	67.00	0.676
	4*	75.75(4.90)	73.50(3.88)	T	0.008	0.016	−2.807	−0.481
Head rotation (°/s)	1*	12.62(5.62)	5.65(2.04)	T	<0.001	<0.001	−7.855	−1.347
	2	15.76(3.75)	12.44(4.32)	W	<0.001	<0.001	71.00	0.664
	3*	15.85(4.00)	15.020(3.98)	T	0.182	0.182	−1.363	−0.234
	4*	30.30(6.98)	29.38(6.89)	T	0.352	0.352	−0.944	−0.162

* indicates that the difference met the normality assumption. In the column of Test, ‘W’ represents Wilcoxon Signed-rank Test (same below).

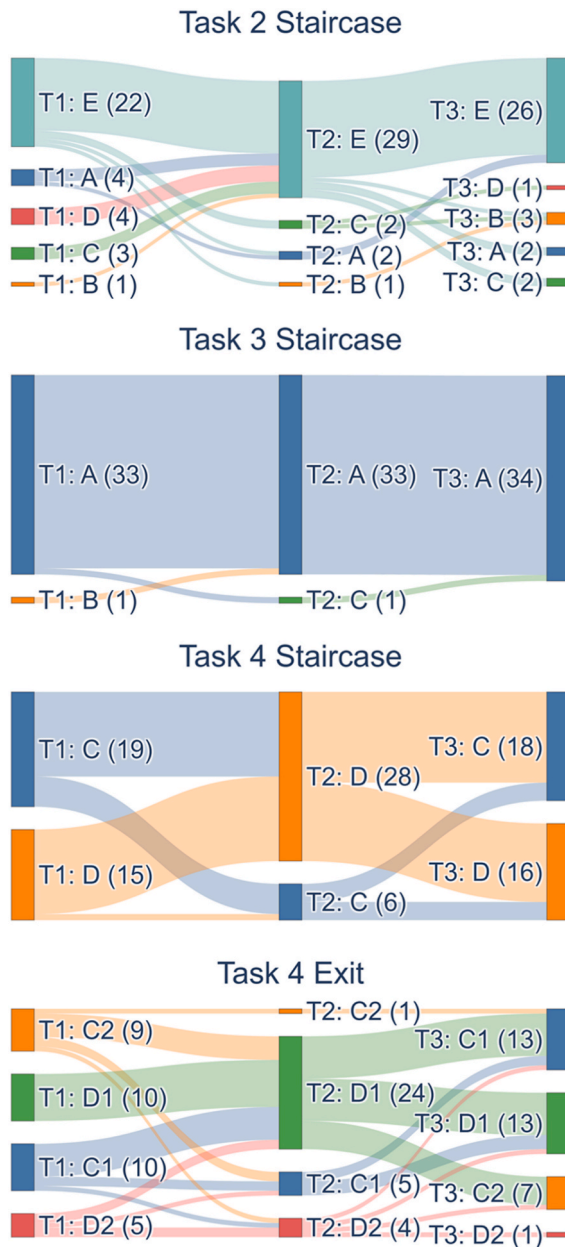


Fig. 8. Descriptive Sankey diagrams showing participants' staircase and exit choices across T1, T2, and T3. All diagrams are based on the 34 participants who completed all three sessions. Numbers in parentheses indicate the number of participants selecting each option at each session, and ribbon widths are proportional to the observed transition counts between adjacent sessions. The diagrams visualize observed choice-transition patterns only; ribbon widths and colors do not indicate statistical significance.

Table 8
Number of participants in T3 whose wayfinding strategies, stair choices, and exit choices were the same as in T1.

Task	Variable	N_same	N_total	Percent_same
1	Strategy	20	34	58.82%
2	Strategy	18	34	52.94%
2	Staircase	17	34	50.00%
3	Strategy	33	34	97.06%
3	Staircase	33	34	97.06%
4	Staircase	21	34	61.76%
4	Exit	13	34	38.24%

changes.

In T3, apart from exit selection in Task 4, more than 50% of participants made the same choices in wayfinding strategy and stair selection in all other tasks as they did in T1, as shown in Table 8. Fig. 9 shows the number of disorientation and detour events among participants in China across the three sessions. In Task 3 and Task 4, no disorientation or detour events were observed in T1, T2, or T3.

3.3.2. Wayfinding performance and head rotation

Statistical results of travel time, travel distance, and head rotation across the three sessions for the 34 participants in China are shown in Table 9. Given the modest sample size and the predominantly non-normal distributions, Friedman tests were used for the T1-T2-T3 comparisons. Significant differences across sessions were found for all outcomes except head rotation in Tasks 3 and 4; therefore, follow-up pairwise comparisons were conducted for the outcomes showing significant omnibus effects.

Follow-up pairwise comparisons were conducted for T2-T3 and T1-T3 (the T1-T2 results are reported in Section 3.2), as shown in Table 10. Compared with T2, performance at T3 was largely maintained: except for Task 1—where travel distance and head rotation were significantly higher after Holm correction. No significant differences were found for the other outcomes. Compared with T1, participants in China at T3 showed significantly shorter travel time and travel distance across all four tasks. For head rotation, values at T3 were significantly lower than at T1 in Tasks 1–3, whereas no significant change was observed in Task 4. Fig. 10 presents a comparison between the Netherlands reference group and China in the three sessions in terms of travel time, travel distance, average speed, and head rotation. Average speed was calculated as travel distance divided by travel time. Because the locomotion speed during active movement in the VR system was fixed, this measure does not represent self-selected walking speed. Instead, it reflects the effective progression rate during the task, including pauses, hesitation, observation, and other wayfinding-related delays.

3.3.3. Summary

Overall, participants in China at T3 performed significantly better than at T1 on most measures, indicating substantial long-term retention of the benefits of prior VR exposure. Relative to T2, evidence of decline was limited mainly to Task 1. In this task, travel distance and head rotation were significantly higher at T3 after Holm correction, whereas the increase in travel time did not remain significant after correction.

3.4. Comparison with the Netherlands reference group after retention

To assess whether long-term familiarization brought participants in China closer to the local reference, performance at T3 was compared with that of the Netherlands reference group. The same between-group procedures used in the baseline comparison were applied here.

3.4.1. Wayfinding decision-making

As shown in Table 11, in Task 1, the Netherlands reference group was more likely to adopt a directional strategy than participants in China at T3. No significant between-group differences in wayfinding strategy were found in Task 2 or Task 3. No significant between-group differences were observed in staircase or exit choice in Task 2, Task 3, or Task 4 (see Table 12). The comparison with the NL1 and NL2 groups indicates that the effect of age can be ruled out to a limited extent, but the effect of familiarity can only be ruled out to a limited extent.

3.4.2. Wayfinding performance and head rotation

As shown in Table 13, participants in China at T3 and the Netherlands reference group exhibited broadly similar wayfinding performance and head-rotation patterns across tasks. In Task 1, participants in China at T3 showed shorter travel time, shorter travel distance, and

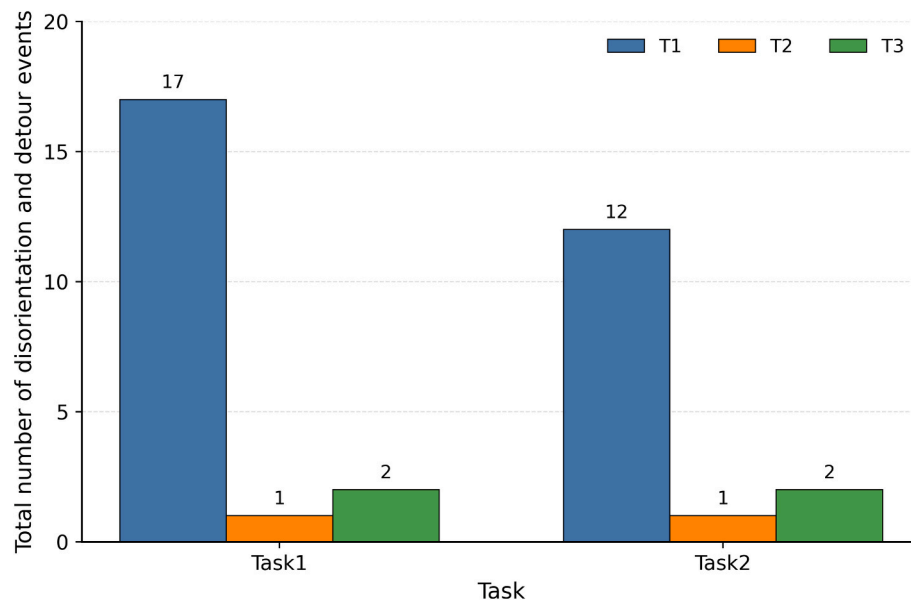


Fig. 9. Total number of disorientation and detour events across T1, T2, and T3.

Table 9

Three-wave comparisons of travel time, travel distance, and head rotation for participants in China (T1, T2, and T3).

Item	Task	T1 Mean(SD)	T2 Mean(SD)	T3 Mean(SD)	p-value	p- holm	Friedman $\chi^2(2)$	Kendall's W
Travel time (s)	1	198.17(44.48)	155.44(15.89)	162.45(17.70)	<0.001	<0.001	25.824	0.38
	2	243.61(44.02)	204.64(33.93)	205.26(15.89)	<0.001	<0.001	32.529	0.48
	3	154.30(30.54)	144.23(27.73)	139.34(19.10)	<0.001	<0.001	15.941	0.23
	4	72.21(13.53)	65.81(17.88)	66.08(17.59)	<0.001	<0.001	17.647	0.26
Travel distance (m)	1	208.80(33.73)	185.77(1.58)	190.03(13.36)	<0.001	<0.001	27.588	0.41
	2	262.30(36.96)	238.17(6.58)	240.95(7.93)	<0.001	<0.001	33.941	0.50
	3	169.94(12.17)	161.80(5.26)	165.82(16.28)	<0.001	<0.001	20.176	0.30
	4	75.75(4.82)	73.50(3.82)	77.94(23.62)	<0.001	<0.001	8.647	0.13
Head rotation (°/s)	1	12.62(5.54)	5.65(2.01)	8.32(3.54)	<0.001	<0.001	27.235	0.40
	2	15.76(3.69)	12.44(4.25)	12.28(3.24)	<0.001	<0.001	25.647	0.38
	3	15.85(3.94)	15.02(3.92)	14.09(3.77)	0.117	0.234	4.294	0.06
	4	30.30(6.87)	29.38(6.79)	28.85(7.23)	0.682	0.682	0.765	0.01

Table 10

Post hoc results for T2-T3 and T1-T3.

Item	T2-T3						T1-T3					
	Task	Test	p- value	p- holm	t or W	effect size	Task	Test	p- value	p- holm	t or W	effect size
Travel time (s)	1	W	0.016	0.064	157.00	0.42	1	W	<0.001	<0.001	56.00	0.71
	2	W	0.489	1.000	257.00	0.12	2	W	<0.001	<0.001	33.00	0.78
	3	W	0.765	1.000	280.00	0.05	3	W	0.002	0.004	119.00	0.52
	4	W	0.739	1.000	278.00	0.06	4	W	0.002	0.004	118.00	0.53
Travel distance (m)	1	W	0.005	0.020	133.00	0.48	1	W	<0.001	<0.001	85.00	0.62
	2	W	0.209	0.627	224.00	0.22	2	W	<0.001	<0.001	49.00	0.73
	3	W	0.952	0.952	294.00	0.01	3	W	0.027	0.050	168.00	0.38
	4	W	0.369	0.738	245.00	0.15	4	W	0.036	0.050	175.00	0.36
Head rotation (°/s)	1	W	0.002	0.008	120.00	0.52	1*	T	<0.001	<0.001	4.40	0.75
	2*	T	0.821	1.000	0.23	0.04	2*	T	<0.001	<0.001	4.79	0.82
	3	W	0.285	0.855	235.00	0.18	3*	T	0.005	0.010	3.01	0.52
	4*	T	0.676	1.000	0.422	0.07	4*	T	0.373	0.373	0.902	0.15

lower head rotation than participants in the Netherlands reference group. However, after Holm correction, only the between-group difference in travel distance remained statistically significant. The detailed statistical results of NL1 and NL2 are provided in Table A. 4 and Table A. 5. After Holm correction, the items that showed significant differences in the comparisons between T3 and NL, T3 and NL1, and T3 and NL2 remained unchanged.

3.4.3. Summary

Overall, after approximately 14 months, the initially observed differences between participants in China and the Netherlands reference group were reduced for most measures. No significant between-group differences were observed in strategy use, staircase choice, or exit choice, except that participants in the Netherlands reference group remained more likely to adopt the directional strategy in Task 1.

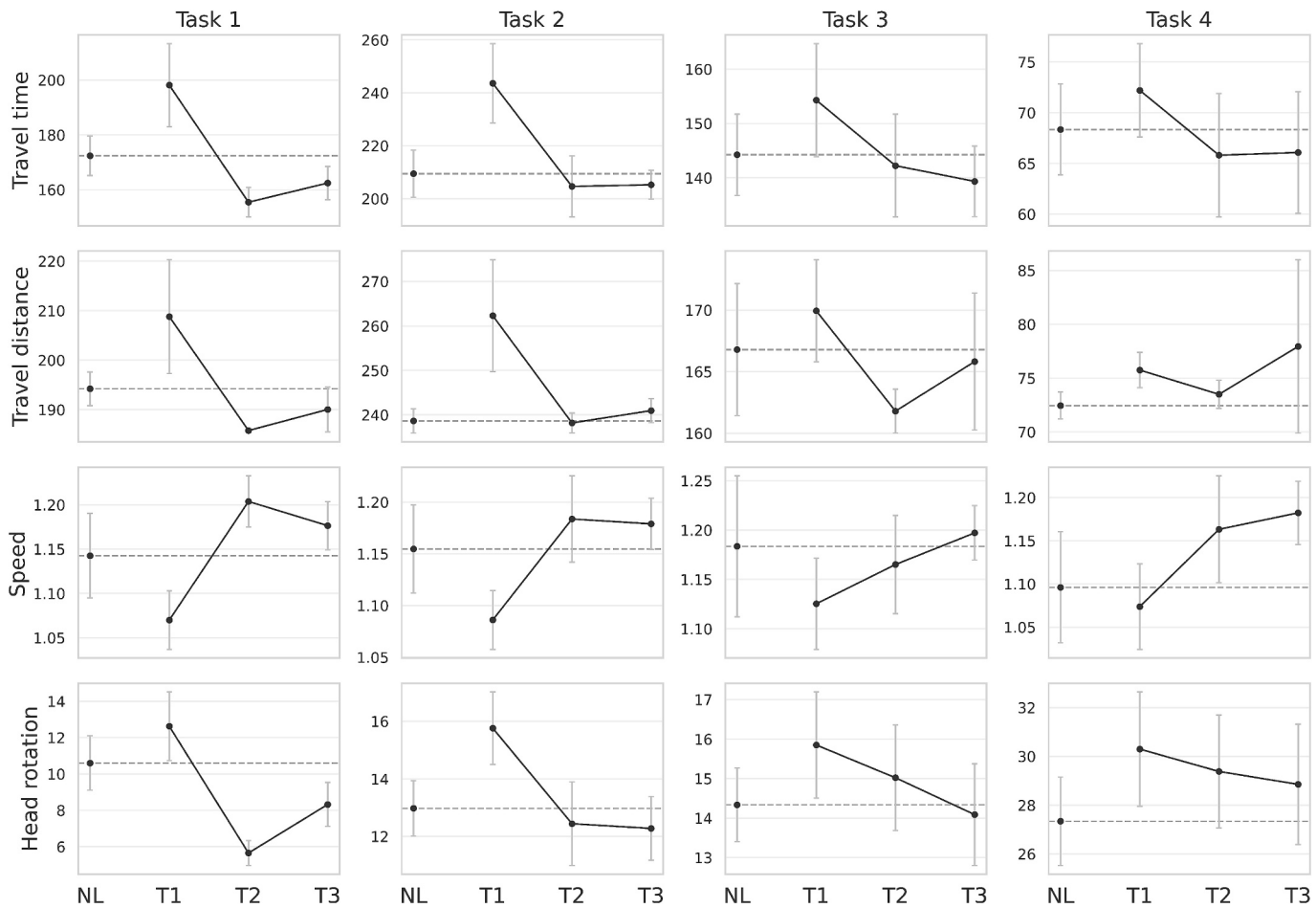


Fig. 10. Descriptive comparison of wayfinding performance between the Netherlands reference group and the China group across sessions. Points represent group or session means, and error bars represent approximate 95% confidence intervals. The dashed horizontal line indicates the Netherlands group mean as a visual reference only. Formal statistical comparisons are reported in the corresponding tables.

Table 11

Test result of wayfinding strategy (T3 vs NL/NL1/NL2).

Group	Task 1	Task 2	Task 3
	Test result (Chi-square)	Test result (Chi-square)	Test result (Fisher exact)
NL vs. T3	$\chi^2(1, N = 70) = 4.67, p = 0.031, \phi = 0.26$	$\chi^2(1, N = 70) = 3.05, p = 0.081, \phi = 0.21$	$p = 0.493$, Cramér's $V = 0.167$
NL1 vs. T3	$\chi^2(1, N = 63) = 3.43, p = 0.064, \phi = 0.23$	$\chi^2(1, N = 63) = 3.24, p = 0.072, \phi = 0.23$	$p = 0.208$, Cramér's $V = 0.196$
NL2 vs. T3	$\chi^2(1, N = 59) = 2.51, p = 0.113, \phi = 0.21$	$\chi^2(1, N = 59) = 2.20, p = 0.138, \phi = -0.24$	Not applicable

Table 12

Test result of staircase and exit choice (T3 vs NL/NL1/NL2).

Group	Task 2 Staircase choice (Fisher exact test)	Task 3 Staircase choice (Fisher exact test)	Task 4 Staircase choice (Chi-square test)	Exit choice (Chi-square test)
NL vs. T3	$p = 0.172$, Cramér's $V = 0.303$	$p = 0.493$, $\phi = 0.167$	$\chi^2(1, N = 70) = 0.89, p = 0.345, \phi = 0.13$	$\chi^2(3, N = 70) = 4.65, p = 0.200$, Cramér's $V = 0.26$
NL1 vs. T3	$p = 0.384$, Cramér's $V = 0.28$	$p = 0.208$, Cramér's $V = 0.20$	$\chi^2(1, N = 63) = 0.84, p = 0.360, \phi = 0.12$	$\chi^2(3, N = 63) = 4.67, p = 0.198$, Cramér's $V = 0.27$
NL2 vs. T3	$p = 0.493$, Cramér's $V = 0.27$	$p = 0.175$, Cramér's $V = 0.22$	$\chi^2(1, N = 59) = 0.46, p = 0.497, \phi = 0.09$	$\chi^2(3, N = 59) = 4.43, p = 0.219$, Cramér's $V = 0.27$

However, because no formal equivalence tests were conducted, the absence of statistically significant differences should not be interpreted as evidence that the two groups were equivalent. For wayfinding performance and head rotation, the two groups also showed largely similar patterns at T3. Taken together, these findings suggest that the performance gap observed at baseline was substantially reduced after VR pre-exposure. Nevertheless, the results should be interpreted as evidence of reduced detectable differences rather than proof that participants in China and the Netherlands became equivalent in their wayfinding

behaviour.

3.5. New wayfinding tasks

Because participants in T1, T2, and T3 completed the same wayfinding tasks, improvement across sessions could reflect memory for specific routes and task order rather than transferable spatial knowledge. To examine this issue, four additional new wayfinding tasks were administered to the VR participants in China (see Section 2.5). Due to

Table 13

Means and standard deviations of wayfinding performance and head rotation for participants in China at T3 and the Netherlands reference group.

Performance	Task	T3 Mean (SD)	NL Mean (SD)	Test	p-value	p-holm	t or U	effect size
Travel time (s)	1	162.45 (17.70)	172.40 (21.85)	U	0.024	0.096	420.00	0.27
	2	205.26 (15.89)*	209.45 (26.86)*	T	0.440	0.782	0.777	0.184
	3	139.34 (19.10)	144.25 (22.60)	U	0.391	0.782	539.00	0.10
	4	66.08 (17.59)	68.35(13.52)	U	0.222	0.666	508.00	0.15
Travel distance (m)	1	190.03 (13.36)	194.22 (10.26)	U	0.007	0.007	384.00	0.32
	2	240.95 (7.93)	238.62 (8.35)	U	0.188	0.564	724.00	0.16
	3	165.82 (16.28)	166.79 (16.20)	U	0.496	0.992	554.00	0.08
	4	77.94 (23.62)	72.45 (3.79)*	U	0.565	0.992	661.00	0.07
Head rotation (°/s)	1	8.32 (3.54)*	10.60 (4.52)	U	0.029	0.116	426.00	0.26
	2	12.28 (3.24)	12.98 (2.92)	U	0.136	0.408	485.00	0.18
	3	14.09 (3.77)	14.34 (2.82)*	U	0.378	0.756	537.00	0.11
	4	28.85 (7.23)	27.34 (5.48)*	U	0.573	0.756	660.00	0.07

Table 14

Comparison of wayfinding performance between New wayfinding Task 1 and T3 Task 3.

Item	New Task 1 Mean (SD)	T3 Task 3 Mean (SD)	Test	p-value	t or W	effect size
Travel time (s)	140.70 (18.45)	139.85 (19.16)	W	0.166	203.00	0.24
Travel distance (m)*	166.98 (15.81)	164.37 (14.21)	T	0.532	0.63	0.11
Head rotation (°/s)*	14.26 (4.53)	14.20 (3.77)	T	0.958	0.05	0.01

scheduling constraints, 33 of the 34 participants completed this follow-up session.

3.5.1. New wayfinding task 1

Table 14 presents participants' performance in new wayfinding Task 1 and T3 Task 3. No significant differences were found in travel time, travel distance, or head rotation. In terms of strategy, 32 participants adopted a floor strategy and 1 adopted a direction strategy. None of the 33 participants simply retraced the original route. These results show that performance was maintained when the route direction was reversed.

3.5.2. New wayfinding task 2

For new wayfinding task 2, the mean travel time was 172.25 s (SD = 8.45), the mean travel distance was 203.80 m (SD = 9.91), and the mean head rotation was 6.54°/s (SD = 4.82). The corresponding BPR values were 1.34 (SD = 0.07), 1.09 (SD = 0.04), and 1.87 (SD = 0.87), respectively. These values were compared with the BPR values of Task 2 in T1 and T3, which was considered the most comparable original task.

As shown in Table 15, New wayfinding Task 2 yielded significantly worse performance than T3 Task 2 on travel time BPR and travel distance BPR, but significantly better performance than T1 Task 2. No significant difference was found for head rotation BPR. Thus, when the destination was changed within the pre-exposed area, performance was lower than in T3 but remained better than in T1.

3.5.3. New wayfinding task 3

Because new wayfinding Task 3 involved an area outside the pre-exposed region, its BPR values were compared only with those of T1 Task 2 (Table 16). All three indices were significantly worse in New Task 3 than in T1 Task 2. This task was more difficult than T1 Task 2, as Exit B had not been previously encountered and could not be inferred from

room numbering. In the post-experiment interview, all participants reported that they had not known of the existence of Exit B or the distribution and naming of the building exits in advance. The benefit of VR pre-exposure did not extend to this unexposed area.

3.5.4. New wayfinding task 4

The photograph used in new wayfinding Task 4 is shown in Fig. 2. In the experimental display, textual annotations were removed and the room number and floor-plan information were blurred. The target location was situated in the central part of the fourth floor and had appeared at least once in each participant's previous pre-exposure routes. Because this task required participants to locate a place from a photograph rather than navigate to a named room or exit, it was not directly compared with the original tasks. Its performance is compared with that of the 2D-map pre-exposure group in Section 3.6.2.

3.5.5. Summary

The new wayfinding tasks were used to examine whether the improvement across T1-T3 reflected more than repeated route learning. New Task 1 showed preserved performance under route reversal. New task 2 showed intermediate performance: worse than T3 Task 2 but better than T1 Task 2. New task 3 showed no advantage for an unexposed area. Overall, the results support transfer within the pre-exposed area but limited transfer beyond it.

3.6. Comparison with the 2D-map pre-exposure group

To further evaluate whether the familiarization benefit was specific to immersive VR pre-exposure, participants' performance was compared with that of an additional 2D-map pre-exposure group. Because this section did not involve cross-national comparisons, and because the main objective was to compare the effects of VR and 2D-map pre-

Table 15

Comparison of BPR values for New Task 2 with T1 Task 2 and T3 Task 2.

Item	T1 Task 2 BPR Mean (SD)	Test	p	t or W	effect size	Item	T3 Task 2 BPR Mean (SD)	Test	p	t or W	effect size
Travel time (s)	1.47 (0.27)	W	0.002	109.00	0.53	travel time (s)*	1.23 (0.09)	T	<0.001	6.08	1.06
Travel distance (m)	1.14 (0.16)	W	0.019	149.00	0.41	travel distance (m)*	1.04 (0.03)	T	<0.001	3.73	0.65
Head rotation (°/s)*	2.15 (0.51)	T	0.117	-1.612	0.28	head rotation (°/s)	1.68 (0.45)	W	0.237	238.00	0.21

Table 16

Comparison of BPR values between New Task 3 and T1 Task 2.

Item	T1 Task 2 BPR Mean (SD)	New Task 3 BPR Mean (SD)	Test	p-value	t or W	effect size
Travel time (s)	1.47 (0.27)	2.24 (1.00)	W	<0.001	72.00	0.65
Travel distance (m)	1.14 (0.16)	2.00 (0.81)	W	<0.001	25.00	0.79
Head rotation (°/s)*	2.15 (0.51)	2.72 (0.76)	T	<0.001	3.66	0.64

Table 17

Means and standard deviations of wayfinding performance and head rotation for participants in T1 and participants in 2DO.

Performance	Task	T1 Mean (SD)	2DO Mean (SD)	Test	p-value	p-holm	t or U	effect size
Travel time (s)	1	198.17 (44.48)	187.25 (23.00)*	U	0.802	1.000	354.00	0.03
	2	243.61 (44.02)	221.97 (13.32)*	U	0.026	0.104	464.00	0.30
	3	154.30 (30.54)	156.52 (25.39)	U	0.531	1.000	305.00	0.08
	4	72.21 (13.53)	74.30 (5.78)*	U	0.048	0.144	229.50	0.27
Travel distance (m)	1	208.80 (33.73)	196.32 (5.96)*	U	0.680	1.000	363.00	0.06
	2	262.30 (36.96)	247.79 (6.15)*	U	0.025	0.100	465.00	0.30
	3	169.94 (12.17)	173.92 (12.79)	U	0.089	0.267	245.00	0.23
	4	75.75 (4.82)	75.37 (5.16)*	U	0.667	1.000	364.00	0.06
Head rotation (°/s)	1	12.62 (5.54)*	9.88 (3.52)*	T	0.033	0.132	-2.19	-0.55
	2	15.76 (3.69)*	14.17 (3.35)	U	0.073	0.219	440.00	0.24
	3	15.85 (3.94)*	16.57 (3.11)*	T	0.495	0.990	0.687	0.19
	4	30.30 (6.87)*	30.14 (7.40)*	T	0.938	0.990	-0.08	-0.02

Table 18

Means and standard deviations of wayfinding performance and head rotation for participants in T3 and participants in 2DO.

Performance	Task	T3 Mean (SD)	2DO Mean (SD)	Test	p-value	p-holm	t or U	effect size
Travel time (s)	1	162.45 (17.70)	187.25 (23.00)*	U	<0.001	<0.001	110.00	0.56
	2	205.26 (15.89)*	221.97 (13.32)*	T	<0.001	<0.001	3.88	1.09
	3	139.34 (19.10)	156.52 (25.39)	U	<0.001	<0.001	153.00	0.46
	4	66.08 (17.59)	74.30 (5.78)*	U	<0.001	<0.001	126.00	0.52
Travel distance (m)	1	190.03 (13.36)	196.32 (5.96)*	U	<0.001	<0.001	115.00	0.55
	2	240.95 (7.93)	247.79 (6.15)*	U	<0.001	<0.001	138.00	0.49
	3	165.82 (16.28)	173.92 (12.79)	U	<0.001	<0.001	145.00	0.47
	4	77.94 (23.62)	75.37 (5.16)*	U	0.275	0.275	279.00	0.15
Head rotation (°/s)	1	8.32 (3.54)*	9.88 (3.52)*	T	0.131	0.262	1.54	0.43
	2	12.28 (3.24)	14.17 (3.35)	U	0.022	0.066	212.00	0.31
	3	14.09 (3.77)	16.57 (3.11)*	U	0.008	0.032	193.00	0.36
	4	28.85 (7.23)	30.14 (7.40)*	U	0.370	0.370	290.00	0.12

exposure, the analysis focused on wayfinding performance and head rotation. Wayfinding decision-making data were not examined in this section.

3.6.1. Original tasks: 2D-map vs. Retained VR familiarization

Because the 2D-map comparator was added after the original longitudinal VR experiment, the interval between map pre-exposure and testing was set to one week. The performance of the 2DO group was compared separately with the VR group in China at T1 and T3. The results are presented in Table 17 and Table 18.

Compared with T1, the 2DO group showed better mean performance in several indices, particularly in Task 1 and Task 2, and the variability of the 2DO group was generally smaller. Before Holm correction, the 2DO group showed lower head rotation in Task 1, shorter travel time and travel distance in Task 2, and longer travel time in Task 4. However, none of these differences remained significant after Holm correction. These results indicate that one week after pre-exposure, 2D-map familiarization produced performance broadly comparable to the initial VR-based wayfinding session, and even showed an advantage over T1 on some indicators before Holm correction, though this advantage was not statistically significant after Holm correction.

In contrast, when the 2DO group was compared with the VR group at T3, the 2DO group performed worse on most indices. The 2DO group had longer travel times in all four tasks and longer travel distances in Tasks 1–3. Most of these differences remained significant after Holm correction. For head rotation, the 2DO group also showed larger values

than the T3 group in Tasks 2 and 3 before correction, with the difference in Task 3 remaining significant after correction. Overall, these results suggest that under the present design, single-session 2D-map pre-exposure did not reach the performance level observed after repeated immersive VR experience and long-term retention.

3.6.2. New transfer tasks: 2D-map vs. VR pre-exposure

The interval between 2D-map pre-exposure and the new wayfinding test was also one week for the 2DN group. To examine the transfer effects of the two pre-exposure methods, the 2DN group was compared with the 33 Chinese participants who completed the new wayfinding tasks after VR pre-exposure. The results are presented in Table 19.

For New Task 1, the VR group significantly outperformed the 2DN group in both travel time and travel distance, whereas no significant difference was found for head rotation. This suggests that immersive VR pre-exposure supported more effective transfer when participants were required to navigate along a reversed route.

For New Task 2, the VR group completed the task significantly faster than the 2DN group. Travel distance did not differ significantly between groups, but head rotation was significantly lower in the VR group. This pattern indicates that when the destination was changed but remained within the pre-exposed area, VR pre-exposure was associated with higher navigation efficiency and lower head-turning activity, possibly reflecting less visual search or route checking. The absence of a significant difference in travel distance further suggests that the 2DN participants did not, on average, become substantially disoriented or take

Table 19

Means and standard deviations of wayfinding performance and head rotation for participants in NT and participants in 2DN.

Item	New task	NT Mean (SD)	2DN Mean (SD)	Test	p-value	p-holm	t or U	effect size
Travel time (s)	1	140.70 (18.45)	191.81 (41.35)	U	<0.001	<0.001	50.00	0.71
	2	172.25 (8.45)	196.55 (20.13)	U	<0.001	<0.001	69.00	0.66
	3	213.56 (95.95)	225.85 (59.19)*	U	0.399	0.399	284.00	0.12
	4	242.49 (37.27)*	329.01 (121.62)*	T	0.007	0.014	3.02	1.06
Travel distance (m)	1	166.98 (15.81)	206.87 (41.50)	U	<0.001	<0.001	106.00	0.56
	2	203.80 (9.91)*	207.44 (11.67)*	T	0.240	0.240	1.189	0.34
	3	262.51 (106.25)	251.48 (67.73)*	U	0.985	0.985	331.00	0.01
	4	275.17 (42.93)	356.53 (125.27)*	U	0.005	0.015	178.00	0.38
Head rotation (°/s)	1	14.26 (4.53)	16.07 (6.79)	U	0.219	0.219	263.00	0.17
	2	6.54 (4.82)	11.07 (3.52)	U	<0.001	<0.001	89.00	0.61
	3	21.74 (6.04)*	16.68 (2.81)*	T	<0.001	<0.001	-4.06	-0.98
	4	16.50 (6.95)	18.88 (5.67)*	U	0.102	0.204	241.00	0.22

markedly more circuitous routes. Rather, they appeared to identify routes of comparable spatial efficiency, but required more time to do so.

For New Task 3, which involved an unexposed area, no significant group differences were found in travel time or travel distance. However, the VR group showed significantly higher head rotation than the 2DN group. This result is consistent with the finding in Section 3.5.3 that neither pre-exposure method provided a clear performance advantage when the target location lay outside the previously experienced area. Although the 2DN group was generally less familiar with the building from an immersive, first-person perspective, its performance was not clearly worse than that of the VR group in this task. One possible explanation is that 2D-map pre-exposure provided participants with a more global overview of the building layout, whereas VR pre-exposure primarily familiarized participants with the areas they had directly explored.

For New Task 4, the VR group performed significantly better than the 2DN group in both travel time and travel distance, while the difference in head rotation was not significant after Holm correction. Because this task required participants to identify a previously seen location from a photograph, the result suggests that immersive VR pre-exposure may support stronger visual-place recognition and spatial transfer than 2D-map pre-exposure.

3.6.3. Summary

The comparisons with the 2D-map pre-exposure group provide further evidence that the familiarization benefits observed in the VR group were not simply attributable to prior access to spatial layout information. For the original tasks, 2D-map pre-exposure produced performance that was broadly comparable to the initial VR test session, but it did not reach the retained performance level observed after repeated VR exposure. This suggests that immersive VR pre-exposure generated more robust familiarization effects for routes and destinations that participants had directly experienced.

The new wayfinding tasks further showed that the advantage of VR pre-exposure varied across task demands, likely reflecting differences in the type of spatial knowledge acquired through immersive versus map-based pre-exposure. VR pre-exposure produced clearer benefits for transfer tasks located within the previously experienced area, particularly when participants needed to navigate a reversed route or identify a destination based on visual scene information. In contrast, when the destination was outside the pre-exposed area, the VR advantage disappeared, and the 2D-map group performed comparably in travel time and travel distance. This pattern suggests that immersive VR pre-exposure may be especially more effective in supporting route-based knowledge, visual-place recognition, and spatial transfer within experienced environments, whereas 2D-map pre-exposure may provide more global but less visually grounded spatial information.

4. Discussion

This study examined whether VR-based pre-exposure can reduce wayfinding difficulties when pedestrians navigate a multi-level building embedded in an unfamiliar national and environmental context, whether repeated exposure reduces the performance gap relative to a local reference group, and whether any gains are retained over the long term. In addition, the study further included new wayfinding tasks and a supplementary 2D floor-plan familiarization condition to examine, respectively, the generalizability of the acquired familiarity and the role of familiarization medium. Overall, the findings suggest that participants in China showed an initial disadvantage when navigating the Dutch university building, that limited VR pre-exposure was followed by substantial short-term and long-term improvements on previously experienced tasks, and that these benefits were only partly transferable to new routes or destinations. The results therefore support the practical value of VR pre-exposure, while also indicating that its effect should be understood primarily as route- and environment-relevant familiarization rather than as the acquisition of complete spatial knowledge.

4.1. Baseline differences in an unfamiliar national and environmental context

At baseline, participants in China at T1 generally required more time, travelled longer distances, and, in several tasks, showed higher head rotation than the Netherlands reference group. This tendency manifests as more observable wayfinding-related behaviours, such as increased head-turning activity, route checking, detours, or repeated turning movements, thereby increasing movement effort and reducing wayfinding efficiency. This interpretation is consistent with prior research showing that head movements can be a valuable indicator for distinguishing familiar from unfamiliar environments in pedestrian wayfinding (Kattenbeck et al., 2025). Previous studies have also shown that familiarity affects visual attention and route choice, and that familiar participants show lower scanning rates during evacuation (Fu et al., 2023).

The above findings may indicate that navigating a building designed and routinely used in a different national, architectural, and institutional context may impose additional cognitive demands. In the absence of prior mental representations, users must rely more heavily on online environmental cues, exploratory scanning, and local decision-making. Nevertheless, head rotation should be interpreted cautiously as an indirect behavioural proxy rather than as direct evidence of underlying cognitive processes. Overall, our findings regarding the effect of familiarity on wayfinding performance are consistent with a broad body of previous work (O'Neill 1992, Hund and Nazarczuk 2009, Al-Alwan and Al-Azzawi 2014, Duran 2016, Muffato and Meneghetti 2020), while extending this line of inquiry to users navigating an environmentally unfamiliar building in a cross-national context.

With regard to wayfinding strategy, our findings are also consistent

with previous evidence that more experienced users tend to prefer a floor strategy over alternative strategies, and that this strategy is generally associated with better wayfinding performance (Hölscher et al., 2006). Previous research has shown that when people know the layout of the environment, they tend to prefer the shortest leg first (SLF) rather than the longest leg first (LLF) strategy (Veerawamy et al., 2011), which is also consistent with our findings. In Tasks 1 and 2 at T1, the central-point strategy and the directional strategy may both be interpreted as LLF-type choices, and these strategies were used more often by participants in China.

It is also noteworthy that the effect of cultural unfamiliarity on wayfinding was more evident in relatively complex tasks. Compared with Tasks 1 and 2, Task 3 involved a shorter and simpler route. One reason is that after completing Task 2 and turning to begin Task 3, participants immediately encountered the staircase in front of them, which substantially reduced navigational uncertainty. This helps explain why the above differences were less apparent in Task 3. Based on this observation, we suggest that future studies on cultural difference and environmental familiarity should preferably employ wayfinding tasks of sufficient complexity to reveal meaningful behavioural variation.

4.2. Short- and long-term improvement following VR pre-exposure

A central finding of this study is the strong short-term improvement from T1 to T2 among participants in China. Across all four tasks, travel time and travel distance decreased significantly after only one prior exposure. Head rotation also decreased significantly in Tasks 1 and 2. Together, these results indicate that even a brief VR experience enabled participants to acquire usable knowledge of the building's route information.

The T3 results further demonstrate that the benefits of VR pre-exposure were not merely transient. At the same time, strategy use remained statistically stable across sessions. Compared with T1, participants in China still showed significantly shorter travel times and shorter travel distances across all tasks approximately 14 months later. Head rotation at T3 also remained lower than at T1 in Tasks 1–3. Although participants may no longer have remembered specific routes in detail after 14 months, the persistently better behavioural performance at T3 suggests that participants retained some degree of functional familiarity with the building. However, because the original T1–T3 measures were behavioural rather than direct measures of spatial knowledge, they cannot by themselves identify the structure of the retained knowledge. The additional new wayfinding tasks were therefore introduced to provide indirect evidence on whether participants could apply the acquired familiarity to new routes and destinations, although this study did not directly measure spatial knowledge structure. Relative to T2, a significant decline at T3 was observed mainly in Task 1. A plausible explanation is that Task 1 was performed at the very beginning of the session, when participants may still have needed time to reactivate their memory of the building layout and environmental details.

Compared with participants in the Netherlands reference group, participants in China at T3 showed smaller differences than those observed at baseline. Most task-specific comparisons were no longer statistically significant, but this finding should be interpreted cautiously because non-significant results do not demonstrate equivalence. Rather, the results indicate that the initial performance gap was reduced within the sensitivity of the present sample and measures. These findings are consistent with the finding that repeated exposure can significantly improve wayfinding performance and reduce psychological stress during evacuation (Lin et al., 2019). Taken together, the present results provide evidence that limited VR pre-exposure can produce durable behavioural benefits and route-relevant familiarity, but the results of new wayfinding tasks indicate that these benefits should not be interpreted as complete or fully transferable spatial knowledge.

The new wayfinding tasks further qualify the interpretation of the long-term effect. Although participants showed durable improvements on previously experienced tasks, their performance on routes or destinations that had not been directly encountered during pre-exposure was relatively limited. This suggests that VR pre-exposure did not necessarily lead to a complete survey-like representation of the entire building. Instead, the acquired familiarity may have been partly route-specific and anchored to previously visited areas, salient decision points, and repeated origin–destination sequences. This finding is important because it prevents an overgeneralized interpretation of VR pre-exposure. The intervention appears to be particularly effective for improving performance on known or structurally similar routes, but less effective when users must infer entirely new paths from incomplete spatial knowledge.

4.3. Route-relevant familiarity and limits of transfer after VR pre-exposure

The present findings suggest that VR pre-exposure can support route-relevant familiarity in a culturally and environmentally unfamiliar building, but they also show that this familiarity is selective rather than fully generalizable. This distinction is important. In practical wayfinding, users do not always need a comprehensive survey representation of the entire building; in many situations, knowing how to reach a small number of critical destinations may already reduce hesitation, inefficient movement, and reliance on real-time decision-making. At the same time, the reduced performance observed in the new wayfinding tasks indicates that VR pre-exposure should not be assumed to prepare users equally well for all possible routes or destinations within a complex building, particularly in previously unexposed areas.

The present findings therefore suggest that VR pre-exposure may be a useful way to improve environmental familiarity before users enter an unfamiliar public building. More importantly, its value may lie not only in exposing users to the building in advance, but also in the specific mode of spatial learning that VR affords. In contrast to purely static representations, VR combines three features that are particularly relevant to wayfinding: active exploration, embodied viewpoint change, and the integration of route-based and configural information. These features may help explain why participants became more efficient on the familiarized tasks, while also clarifying why the benefits did not fully generalize to all new routes. In a VR environment, users are not merely told where destinations are located; they experience the building by moving through it, making directional choices, and adjusting their behaviour in response to environmental cues. This process may be especially useful for users who are unfamiliar with the building and its wayfinding logic, because they cannot rely fully on prior expectations about the organization of the environment and therefore need opportunities to construct route and spatial knowledge through direct interaction. More frequent interaction with the environment has been associated with more structured knowledge organization in memory (Al-Alwan and Al-Azzawi 2014). By allowing users to scan, turn, and progress through the space in a way that resembles actual movement, VR may support the formation of behaviourally usable route and spatial knowledge that can be applied during subsequent wayfinding.

In addition, VR may support both route-based learning and partial configural learning, but the present results suggest that these two forms of knowledge were not acquired to the same extent. Prior research has suggested a trade-off between these two forms of spatial learning: comprehensive configural information can support the acquisition of an overall spatial structure but may come at the cost of accurate route-following, whereas route-based presentation may support immediate wayfinding performance while limiting broader configural understanding (Münzer et al., 2012). In the present study, repeated exposure to specific tasks likely strengthened memory for routes, landmarks, vertical transitions, and decision points. This interpretation is consistent with the reductions in travel time, travel distance, and head rotation on

repeated tasks: participants became more efficient on familiarized routes and appeared to require less visual search and orientation checking during navigation.

However, the reduced performance in the new wayfinding tasks indicates that participants may not have developed a complete configurational understanding of the building. VR pre-exposure appeared to support transfer more clearly when the new task remained within the previously experienced area or could be solved by relying on remembered route segments, visual scenes, or reversed route knowledge. In contrast, when the destination was outside the pre-exposed area, the benefit of VR pre-exposure was less apparent. This pattern suggests that VR may reduce, but not eliminate, the trade-off between route-based and survey-based learning. In other words, immersive pre-exposure can help users become familiar with experienced routes and areas, but this familiarity does not necessarily generalize to all unfamiliar destinations within the same building.

These considerations also help explain why VR pre-exposure may be particularly useful in complex environments and for users with relatively limited spatial ability. Previous research has shown that floor-plan complexity negatively affects wayfinding performance (O'Neill 1992). At the same time, map-based learning often depends more heavily on individual visuo-spatial ability, and inter-individual differences tend to become more pronounced after map learning than after video-based learning (Muffato 2016). It has also been reported that participants with a poor sense of direction may derive limited benefit from simple maps (Stites et al., 2020). In this context, VR may offer a relatively accessible familiarization method because it reduces the need to translate abstract symbolic information into movement decisions. Instead, users can learn through perceptually grounded interaction with the environment itself. Nevertheless, the present findings also indicate that the effectiveness of VR familiarization depends on the match between the pre-exposed routes and the later wayfinding demands. If flexible navigation across a wider building area is required, VR training may need to include broader exploration, multiple route alternatives, and 2D map support for understanding the overall spatial structure.

Finally, the present findings suggest that the value of VR pre-exposure may not be limited to performance improvement alone but may also include practical applicability across user groups. User-experience measures showed no significant differences between participants in China and the Netherlands, indicating that the WayR software was experienced similarly across the two participant groups. This does not by itself demonstrate universal cross-cultural applicability, but it suggests that well-designed VR-based familiarization tools can be acceptable to users from different national backgrounds and may serve as a practical method for preparing diverse visitors to navigate unfamiliar public buildings.

4.4. VR pre-exposure compared with 2D floor-plan familiarization

The supplementary 2D floor-plan familiarization condition provides an exploratory benchmark for interpreting whether the benefits observed in the VR group were merely attributable to prior access to spatial layout information. Floor plans are a common and efficient way to convey building configuration, but they require users to translate abstract, allocentric layout information into egocentric movement decisions during actual navigation. VR pre-exposure, in contrast, presents the environment from a first-person perspective and allows users to experience sequential decision points, visual scenes, turns, and vertical transitions in a manner closer to real navigation. The comparison between these two familiarization methods therefore helps clarify not only whether pre-exposure is beneficial, but also what type of spatial knowledge different media may support.

For the original wayfinding tasks, participants who received 2D-map pre-exposure performed broadly comparably to the initial VR test session, but their performance did not reach the retained level observed after repeated VR exposure. This pattern suggests that access to layout

information alone may not have been sufficient to reproduce the full benefit observed in the longitudinal VR group. In particular, repeated VR exposure appeared to support more robust familiarity with routes and destinations that participants had directly experienced. However, this interpretation should be made cautiously because the 2D-map condition was added as a supplementary comparison and was not part of the original longitudinal randomized design. Differences between the two conditions may therefore reflect not only the familiarization medium, but also differences in exposure format, learning strategy, or repetition.

The new wayfinding tasks further showed that the advantage of VR pre-exposure varied across task demands. VR pre-exposure produced clearer benefits when the transfer task was located within the previously experienced area, especially when participants needed to navigate a reversed route or identify a destination based on visual scene information. These results are consistent with the idea that immersive VR supports route-based knowledge, recognition of visually experienced places, and transfer within familiarized areas (Sauzéon et al., 2016, Afrooz et al., 2018, He et al., 2019, Sun et al., 2019, Kim and Bock 2021). In such tasks, participants could rely on remembered scenes, decision points, route segments, and vertical transitions encountered during pre-exposure, even when the exact route direction or task requirement differed from the original tasks.

By contrast, when the destination was located outside the pre-exposed area, the advantage of VR pre-exposure became less apparent, and the 2D-map group performed comparably in travel time and travel distance. This finding is important because it refines the interpretation of VR-based learning. If VR pre-exposure had produced a complete survey-like representation of the building, a stronger advantage might have been expected even for destinations outside the directly experienced area. Instead, the results suggest that VR-based familiarization was strongest within the experienced spatial envelope and did not necessarily generalize to unvisited regions. In this respect, 2D floor plans may provide more global but less visually grounded spatial information, whereas VR may provide richer egocentric and visual-route information but with transfer benefits that depend on the scope of prior exploration.

Overall, the comparison with 2D floor-plan familiarization suggests that VR pre-exposure should not be understood simply as another way of presenting layout information. Its main advantage appears to lie in supporting visually grounded, route-relevant familiarity with previously experienced areas. At the same time, the absence of a consistent VR advantage for destinations outside the pre-exposed area indicates that immersive familiarization alone may be insufficient for acquiring complete building-wide spatial knowledge.

4.5. Implications for safety science and building management

The present findings have several implications for the safety science domain. First, they reinforce the importance of considering user diversity in safety evaluation. Public buildings are often assessed as if users were homogeneous, yet our results suggest that performance can differ substantially when users enter an environment shaped by unfamiliar cultural and spatial conventions. In evacuation and emergency response, even modest inefficiencies or hesitation may influence congestion, egress utilization, and risk exposure.

More importantly, the findings indicate that VR-based pre-exposure may serve as a feasible supplementary risk-mitigation tool for unfamiliar populations. It should not be viewed as a substitute for inclusive architectural design, clear signage, or safety-oriented building management. Rather, it may complement these measures, especially when physical modification is costly, when users are temporary visitors, or when large numbers of people need to familiarize themselves with a facility before arrival. However, signs, alarms, and movable facilities should still be designed with cross-cultural considerations (Kaminka and Fridman 2018, Jiayu et al., 2025), because unfamiliar users rely more heavily on environmental information. VR-based pre-exposure allows

temporary or foreign users to acquire environmental knowledge before actual use, and the present findings suggest that such pre-exposure may remain beneficial even when delivered months in advance of actual arrival.

From a safety-management perspective, this is particularly promising for internationalized facilities such as airports, hospitals, university buildings, and conference venues, where a large proportion of users may have limited time to become familiar with the space naturally. Importantly, the practical value of VR pre-exposure does not depend on users acquiring complete spatial knowledge of an entire building. In many real-world situations, partial but reliable familiarity with a small number of critical routes may already be valuable. For example, international travellers arriving at a large airport may need to transfer quickly to intercity rail, metro services, or long-distance coaches. In such cases, pre-exposure several days or weeks before departure, or offline mobile pre-exposure during the flight, could help users become familiar with the route from passport control or baggage claim to the relevant transport platform. Even if users do not acquire a full survey representation of the airport, prior exposure to key decision points, vertical transitions, and route sequences may reduce hesitation and improve time-critical wayfinding after arrival. This may be particularly relevant because transport buildings and public facilities often differ substantially across countries. For instance, large high-speed railway stations in China commonly involve security screening, large waiting halls, ticket-gate access before platform entry, and a relatively clear separation from surrounding urban commercial buildings. By contrast, many railway stations in Japan are more compact, more directly integrated with commercial complexes, and typically do not involve the same sequence of security checks and waiting-hall procedures. Such differences mean that international users may not only face an unfamiliar layout, but also unfamiliar operational logic. VR-based pre-exposure could therefore help users understand both the spatial sequence and the procedural structure of an unfamiliar facility before they arrive.

With the development of virtual reality, digital twins, and mobile 3D applications, it is now easier than before to deploy pre-exposure programs on personal devices such as smartphones and laptops (i.e. desktop VR), allowing external visitors to familiarize themselves with a building in advance. In addition, emerging technologies such as Neural Radiance Fields (NeRF) and 3D Gaussian Splatting (3DGS) make it increasingly feasible to obtain highly realistic 3D environmental models rapidly and at relatively low cost. This is beneficial not only for VR developers but also potentially for users, as higher realism may improve the effectiveness of pre-exposure.

4.6. Limitations

Several limitations should be acknowledged. First, the between-group comparisons cannot be attributed solely to culture. Although the study was designed as a cross-national comparison between China and the Netherlands, cultural constructs were not directly measured. The observed differences may reflect country-related background, local environmental conventions, age, education, prior building experience, and building-specific unfamiliarity. Thus, the baseline contrast should be interpreted as a cross-group comparison under cultural and environmental unfamiliarity, rather than as a direct test of specific cultural mechanisms. In addition, the design was asymmetric: the China group completed repeated sessions, whereas the Netherlands reference group completed only one session. Therefore, the China T3 versus NL comparison should not be interpreted as a longitudinal test of convergence, equivalence, or normalization. The reduced between-group differences may partly reflect repeated testing, procedural familiarity, ceiling effects, or regression to the mean. Because the Netherlands group was not retested over equivalent intervals, these alternatives cannot be ruled out, and the Netherlands group should be regarded as a descriptive local-reference benchmark rather than a longitudinal control group.

Second, the sample consisted mainly of university students, and the

China participants were recruited from one university with a relatively narrow age range. Students may have relatively strong memory, attention, and spatial-learning abilities, which may partly contribute to the observed retention effect. More heterogeneous user groups should be examined, especially given evidence that active navigation in VR can affect episodic memory in older adults (Sauzéon et al., 2016).

Third, the study used a single university teaching building. Although it was multi-level and suitable for experimental control, its layout was relatively regular and the tasks were comparatively simple. VR pre-exposure may be less durable or transferable in more complex public facilities, where irregular layouts, high pedestrian density, temporary obstacles, dynamic signage, and operational changes may affect route choice and orientation.

Fourth, although previous studies suggest that pedestrian wayfinding performance in VR can be broadly comparable to that in real environments (Coutrot et al., 2019, Dong et al., 2022, van Beek et al., 2024), VR is not identical to real-world navigation. The locomotion interface, sensory richness, social interaction, time pressure, and emergency stress were simplified. Therefore, implications for actual wayfinding and evacuation behaviour should be interpreted with caution, particularly in crowded public buildings.

Fifth, the longitudinal comparison should not be interpreted as isolating a VR-specific causal effect. Because no repeated-testing control group without pre-exposure was included, part of the improvement may reflect task repetition, procedural familiarity, or memory of the experimental setting. The supplementary 2D floor-plan experiment provides a useful benchmark, but it was added after the original longitudinal study, involved a different retention interval, and may have differed in motivation, exposure experience, and learning strategy. Therefore, the VR-2D comparison should be interpreted as exploratory.

Finally, head rotation was used only as an indirect and coarse indicator of orientation-related exploratory behaviour, not as a direct measure of gaze allocation or visual attention. Although reduced head rotation may indicate lower visual search demands or greater route familiarity, it cannot reveal which environmental features participants attended to or how visual information was processed at decision points.

4.7. Future work

Future research should proceed in three directions. First, studies should recruit more representative and better matched samples, more strictly control participants' prior familiarity with the environment, and consider individual factors such as handedness and spatial ability. This would allow between-group differences to be interpreted more cautiously in relation to national or cultural background, and would enable formal equivalence or non-inferiority testing to determine whether post-exposure performance falls within a predefined practically equivalent range relative to a local reference group. More advanced data-collection devices, such as eye trackers, should also be used to capture pause duration, location-specific hesitation, fixation on signage, and physiological indicators of stress.

Second, VR pre-exposure should be directly compared with other familiarization methods, such as 2D floor plans, videos, guided tours, and mobile navigation aids. Fully controlled experiments are needed to evaluate the relative effectiveness and cost-efficiency of these methods under equivalent exposure duration, task design, sample size, and retention intervals.

Third, future studies should extend the research to more diverse and safety-critical building types, such as hospitals, metro stations, airports, and high-rise public buildings. Scenario and task design should also be refined by introducing obstacles, dynamic signage, crowding, social interaction, environmental stressors, and changing access conditions, so that virtual emergency scenarios more closely approximate real safety-critical situations.

5. Conclusion

This study investigated whether VR-based pre-exposure can reduce the initial wayfinding disadvantage observed when users navigate a multi-level building in an unfamiliar national and environmental context, whether repeated exposure makes behaviour more similar to that of a local reference group, and whether these benefits persist over time. Participants in China initially showed lower wayfinding efficiency than the Netherlands reference group, but VR pre-exposure was followed by significant improvements in travel time, travel distance, and, in some tasks, head rotation. Importantly, a substantial part of this improvement remained evident approximately 14 months later.

The additional new wayfinding tasks show that these benefits should not be interpreted as complete or fully generalizable spatial knowledge. Rather, VR pre-exposure appears to support durable route-relevant familiarity, especially for previously experienced routes, destinations, and decision points. The supplementary 2D floor-plan condition further provides exploratory evidence for comparing VR with a conventional familiarization method, although this comparison requires confirmation in a fully controlled randomized design.

From a practical perspective, the findings suggest that VR pre-exposure may be useful even when it does not produce a comprehensive mental representation of an entire building. In many public facilities, such as airports, railway stations, hospitals, universities, and conference venues, unfamiliar users may only need to become familiar with a limited number of critical routes before arrival. VR-based pre-arrival familiarization may therefore serve as a complementary safety-management tool for reducing hesitation, detours, and inefficient route choice among diverse user populations. However, further research in more complex, dynamic, and safety-critical environments is needed before the long-term effectiveness and transferability of VR pre-exposure can be generalized to real-world public buildings.

Appendix A

Table A1

Means and standard deviations of wayfinding performance and head rotation for participants in China at T1 and participants in NL1 group (under 30 group, 29 participants).

Item	Task	NL1 Mean(SD)	T1 Mean (SD)	Test	p-value	p-holm	t or U	effect size
Travel time (s)	1	170.48 (16.05)*	198.17 (44.48)	U	0.011	0.033	677.00	0.32
	2	213.03 (28.36)	243.61 (44.02)	U	<0.001	0.001	779.00	0.50
	3	144.89 (23.98)	154.30 (30.54)	U	0.071	0.144	624.00	0.23
	4	69.43 (13.63)	72.21 (13.53)	U	0.363	0.363	559.00	0.11
Travel distance (m)	1	194.44 (10.46)	208.80 (33.73)	U	0.148	0.148	598.00	0.18
	2	237.59 (4.41)*	262.30 (36.96)	U	<0.001	<0.001	890.00	0.69
	3	168.08 (17.60)	169.94 (12.17)	U	0.033	0.066	648.00	0.27
	4	72.53 (3.48)*	75.75 (4.82)	U	0.005	0.015	695.00	0.35
Head rotation (°/s)	1	10.90 (4.77)*	12.62 (5.54)*	T	0.203	0.220	−1.29	−0.33
	2	13.28 (3.05)	15.76 (3.69)*	U	0.003	0.012	707.00	0.37
	3	14.38 (3.01)*	15.85 (3.94)*	T	0.110	0.220	−1.62	−0.41
	4	27.09 (5.17)*	30.30 (6.87)*	T	0.047	0.141	−2.03	−0.51

Table A2

Means and standard deviations of wayfinding performance and head rotation for participants in China at T1 and participants in NL2 group (less familiar group, 25 participants).

Item	Task	NL2 Mean(SD)	T1 Mean (SD)	Test	p-value	p-holm	t or U	effect size
Travel time (s)	1	175.33 (23.28)	198.17 (44.48)	U	0.050	0.150	551.00	0.25
	2	215.10 (28.73)	243.61 (44.02)	U	<0.001	0.002	661.00	0.47
	3	148.73 (22.89)	154.30 (30.54)	U	0.374	0.748	483.00	0.12

(continued on next page)

CRedit authorship contribution statement

Chenyi Yang: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Dewei Li:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition. **Zhicheng Dai:** Writing – review & editing, Validation, Methodology. **Yan Feng:** Writing – review & editing, Visualization, Software, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement and funding

The author gratefully acknowledges Professor Yan Feng for her generous support. This cross-cultural study would not have been possible without the data, software, funding, and invaluable academic guidance she provided throughout the research process. The author also thanks Professor Dewei Li for providing laboratory space and equipment, and for his insightful comments on the methodological issues and limitations of this work. Sincere thanks are extended to Dr. Zhicheng Dai, whose constructive suggestions substantially improved the experimental design and who offered crucial assistance in organizing and conducting the experiments. The author is also grateful to the 70 participants from the two countries for their time and commitment to this study.

This work was supported by the Seed Funding project “Comparison of pedestrian wayfinding and evacuation behavior across cultures”.

Table A2 (continued)

Item	Task	NL2 Mean(SD)	T1 Mean (SD)	Test	p-value	p-holm	t or U	effect size
Travel distance (m)	4	69.90 (14.47)	72.21 (13.53)	U	0.399	0.748	480.00	0.11
	1	193.22 (7.35)	208.80 (33.73)	U	0.133	0.133	523.00	0.20
	2	239.18 (9.46)	262.30 (36.96)	U	<0.001	0.002	747.00	0.64
	3	166.79 (16.88)	169.94 (12.17)	U	0.026	0.050	570.00	0.29
Head rotation (°/s)	4	72.24 (3.87)*	75.75 (4.82)	U	0.002	0.006	623.00	0.40
	1	10.34 (3.81)*	12.62 (5.54)*	T	0.071	0.142	−1.84	−0.46
	2	12.65 (1.96)*	15.76 (3.69)*	T	<0.001	0.002	−4.11	−0.99
	3	14.52 (2.83)*	15.85 (3.94)*	T	0.163	0.163	−1.41	−0.37
	4	26.79 (4.38)*	30.30 (6.87)*	T	0.032	0.096	−2.20	−0.58

Table A3

Test results of the user experience questionnaire (NL vs. T1).

Item	T1	NL	Test	p-value	p-holm	t or U	effect size
Face validity (1–5)							
Total score of face validity	3.94 (0.57)	3.92 (0.48)*	U	0.598	1.000	568.00	0.06
Realism of the virtual building	4.09 (0.45)	4.11 (0.74)	U	0.831	1.000	628.00	0.03
Realism of the virtual furniture	3.88 (0.98)	3.86 (0.71)	U	0.618	1.000	573.00	0.06
Realism of the movement abilities	3.12 (1.00)	3.05 (0.85)	U	0.614	1.000	572.00	0.06
Realism of the evacuation alarm sound	4.68 (0.54)	4.64 (0.75)	U	0.672	1.000	639.50	0.05
SSQ (0–3)							
Total score	13.85 (15.96)	14.05 (13.05)	U	0.466	1.000	674.00	0.09
Nausea	17.40 (24.46)	15.10 (12.98)	U	0.494	1.000	668.50	0.08
Oculomotor	20.29 (21.17)	19.82 (21.03)	U	0.668	1.000	648.00	0.05
Disorientation	34.80 (41.58)	35.96 (38.28)	U	0.495	1.000	669.00	0.08
PQ (1–7)							
Involvement	5.15 (0.94)*	4.89 (0.84)*	T	0.218	0.436	1.25	0.30
Sensory fidelity	5.50 (0.96)*	5.02 (0.88)*	T	0.035	0.105	2.15	0.51
Immersion	5.61 (0.59)*	5.65 (0.63)*	T	0.802	0.802	−0.25	−0.06
Interface quality	5.42 (0.76)*	3.20 (0.90)*	T	<0.001	<0.001	11.05	2.64
SUS (1–5)							
I would like to use the system frequently	3.38 (1.23)	3.61 (1.14)	U	0.406	1.000	680.50	0.10
I found the system unnecessarily complex	2.38 (1.37)	1.53 (0.64)	U	0.007	0.070	396.50	0.32
I thought the system was easy to use	4.12 (0.84)	4.42 (0.64)	U	0.145	1.000	726.00	0.17
I think that I would need the support of a technical person to be able to use this system	2.35 (1.32)	1.89 (1.10)	U	0.118	0.944	485.50	0.19
I found the various functions in this system were well integrated	3.85 (0.96)	3.86 (0.67)	U	0.866	1.000	598.50	0.02
I thought there was too much inconsistency in this system	2.15 (0.99)	1.67 (0.71)	U	0.029	0.261	439.50	0.26
I would imagine that most people would learn to use this system very quickly	4.38 (0.82)	4.22 (0.85)	U	0.356	1.000	541.00	0.11
I found the system very cumbersome to use	1.76 (1.02)	1.72 (1.02)	U	0.746	1.000	587.00	0.04
I felt very confident using the system	4.24 (0.99)	4.17 (0.87)	U	0.564	1.000	566.50	0.07
I needed to learn a lot of things before I could get going with this system	1.65(0.95)	1.56 (0.96)	U	0.468	1.000	558.50	0.09

Table A4

Means and standard deviations of wayfinding performance and head rotation for participants in China at T3 and participants in NL1 group.

Item	Task	NL1 Mean (SD)	T3 Mean (SD)	Test	p-value	p-holm	t or U	effect size
Travel time (s)	1	170.48 (16.05)*	162.45 (17.70)	U	0.023	0.092	328.00	0.29
	2	213.03 (28.36)	205.26 (15.89)*	U	0.356	0.712	426.00	0.12
	3	144.89 (23.98)	139.34 (19.10)	U	0.440	0.712	437.00	0.10
	4	69.43 (13.63)	66.08 (17.59)	U	0.107	0.321	376.00	0.20
Travel Distance (m)	1	194.44 (10.46)	190.03 (13.36)	U	0.006	0.024	295.00	0.34
	2	237.59 (4.41)*	240.95 (7.93)	U	0.140	0.420	600.00	0.19
	3	168.08 (17.60)	165.82 (16.28)	U	0.408	0.816	433.00	0.10
	4	72.53 (3.48)*	77.94 (23.62)	U	0.720	0.816	519.00	0.05
Head rotation (°/s)	1	10.90 (4.77)*	8.32 (3.54)*	T	0.018	0.072	2.42	0.61
	2	13.28 (3.05)	12.28 (3.24)	U	0.085	0.255	368.00	0.22
	3	14.38 (3.01)*	14.09 (3.77)	U	0.456	0.912	439.00	0.09
	4	27.09 (5.17)*	28.85 (7.23)	U	0.544	0.912	537.00	0.08

Table A5

Means and standard deviations of wayfinding performance and head rotation for participants in China at T3 and participants in NL2 group.

Item	Task	NL2 Mean (SD)	T3 Mean (SD)	Test	p-value	p-holm	t or U	effect size
Travel time (s)	1	175.33 (23.28)	162.45 (17.70)	U	0.014	0.056	264.00	0.32
	2	215.10 (28.73)	205.26 (15.89)*	U	0.198	0.214	341.00	0.17
	3	148.73 (22.89)	139.34 (19.10)	U	0.050	0.150	297.00	0.26
	4	69.90 (14.47)	66.08 (17.59)	U	0.107	0.214	320.00	0.21
Travel Distance (m)	1	193.22 (7.35)	190.03 (13.36)	U	0.008	0.032	252.00	0.35
	2	239.18 (9.46)	240.95 (7.93)	U	0.198	0.594	509.00	0.17
	3	166.79 (16.88)	165.82 (16.28)	U	0.500	0.850	381.00	0.09
	4	72.24 (3.87)*	77.94 (23.62)	U	0.425	0.850	477.00	0.10
Head rotation (°/s)	1	10.34 (3.81)*	8.32 (3.54)*	T	0.047	0.188	2.04	0.54
	2	12.65 (1.96)*	12.28 (3.24)	U	0.182	0.546	338.00	0.17
	3	14.52 (2.83)*	14.09 (3.77)	U	0.357	0.714	365.00	0.12
	4	26.79 (4.38)*	28.85 (7.23)	U	0.416	0.714	478.00	0.11

Table A6

Demographic characteristics of the participants in VR and 2D map pre-exposure group.

Descriptive information	Category	VR pre-exposure	2D pre-exposure	Test result
Age		23.44 (1.31)	23.13 (1.75)	Mann-Whitney U = 779.50, p = 0.268, r = 0.13
Gender	Male	18, 52.9%	20, 50%	p = 0.801 $\chi^2(1, N = 74) = 0.06$, $\phi = 0.03$
	Female	16, 47.1%	20, 50%	
Highest education level	High school or equivalent	0, 0.0%	0, 0.0%	p = 0.894 Cramér's V = 0.073
	Bachelor's degree or equivalent	29, 85.3%	32, 80.0%	
	Master's degree or equivalent	4, 11.8%	6, 15.0%	
	Doctoral degree or equivalent	1, 2.9%	2, 5.0%	
Previous experience with VR	Never	6, 17.6%	4, 10.0%	p = 0.781 Cramér's V = 0.121
	Seldom	18, 52.9%	22, 55.0%	
	Sometimes	9, 26.5%	12, 30.0%	
	Often	1, 2.9%	2, 5.0%	
Familiarity with any computer gaming	Not at all familiar	5, 14.7%	5, 12.5%	p = 0.932 Cramér's V = 0.107
	little familiar	7, 20.5%	9, 22.5%	
	Moderately familiar	9, 26.5%	11, 27.5%	
	Quite-a-bit familiar	9, 26.5%	8, 20.0%	
	Very familiar	4, 11.8%	7, 17.5%	
Frequency of Emergency drill	Never	1, 2.9%	1, 2.5%	p = 0.942 Cramér's V = 0.072
	Seldom	10, 29.4%	10, 27.5%	
	Sometimes	20, 58.9%	24, 60.0%	
	Often	3, 8.8%	5, 10.0%	
Familiarity with the faculty building	Not at all familiar	34, 100.0%	40, 100%	---
	little familiar	0, 0.0%	0, 0.0%	
	Moderately familiar	0, 0.0%	0, 0.0%	
	Quite-a-bit familiar	0, 0.0%	0, 0.0%	
	Very familiar	0, 0.0%	0, 0.0%	

Table A7

Demographic characteristics of the participants in 2DO and 2DN.

Descriptive information	Category	2DO	2DN	Test result
Age		23.00 (1.45)	23.25 (2.00)	Mann-Whitney U = 185.00, p = 0.698, r = 0.11
Gender	Male	8, 40.0%	12, 60.0%	p = 0.206 $\chi^2(1, N = 40) = 1.60$, $\phi = 0.20$
	Female	12, 60.0%	8, 40.0%	
Highest education level	High school or equivalent	0, 0.0%	0, 0.0%	p = 0.826 Cramér's V = 0.141
	Bachelor's degree or equivalent	17, 85.0%	15, 75.0%	
	Master's degree or equivalent	2, 10.0%	4, 20.0%	
	Doctoral degree or equivalent	1, 5.0%	1, 5.0%	
Previous experience with VR	Never	1, 5.0%	3, 5.0%	p = 0.421 Cramér's V = 0.296
	Seldom	12, 60.0%	10, 50.0%	
	Sometimes	5, 25.0%	7, 30.0%	
	Often	2, 10.0%	0, 0.0%	
Familiarity with any computer gaming	Not at all familiar	2, 10.0%	3, 15.0%	p = 0.702 Cramér's V = 0.246
	little familiar	5, 25.0%	4, 20.0%	
	Moderately familiar	4, 20.0%	7, 35.0%	
	Quite-a-bit familiar	4, 20.0%	4, 20.0%	
	Very familiar	5, 25.0%	2, 10.0%	

(continued on next page)

Table A7 (continued)

Descriptive information	Category	2DO	2DN	Test result
Frequency of Emergency drill	Never	1, 5.0%	0, 0.0%	p = 0.727 Cramér's V = 0.210
	Seldom	6, 30.0%	4, 20.0%	
	Sometimes	11, 55.0%	13, 65.0%	
	Often	2, 10.0%	3, 15.0%	
Familiarity with the faculty building	Not at all familiar	20, 100.0%	20, 100%	---
	little familiar	0, 0.0%	0, 0.0%	
	Moderately familiar	0, 0.0%	0, 0.0%	
	Quite-a-bit familiar	0, 0.0%	0, 0.0%	
	Very familiar	0, 0.0%	0, 0.0%	

Data availability

The authors do not have permission to share data.

References

Acela, L.M.B., et al., 2025. Pedestrian crossing behavior in social groups: exploring cultural contexts through a comparative study. *Case Studies on Transport Policy* 19, 101333.

Afroz, A., et al., 2018. Effects of active and passive exploration of the built environment on memory during wayfinding. *Appl. Geogr.* 101, 68–74.

Al-Alwan, H., Al-Azzawi, A., 2014. Wayfinding performance in the built environment. *Emirates J Eng Res (EJER)* 19 (2), 1–17.

Barrett, T.J., Hegarty, M., 2016. Effects of interface and spatial ability on manipulation of virtual models in a STEM domain. *Comput. Hum. Behav.* 65, 220–231.

Bitgood, S., et al., 2013. Pedestrian choice behavior at shopping mall intersections in China and the United States. *Environ. Behav.* 45 (8), 1019–1032.

Brooke, J., 1996. SUS-A quick and dirty usability scale. *Usability Evaluation in Industry* 189 (194), 4–7.

Chen, J., et al., 2023. Cross-cultural assessment of the effect of spatial information on firefighters' wayfinding performance: a virtual reality-based study. *Int. J. Disaster Risk Reduct.* 84, 103486.

Coutrot, A., et al., 2019. Virtual navigation tested on a mobile app is predictive of real-world wayfinding navigation performance. *PLoS One* 14 (3), e0213272.

Dong, W., et al., 2022. Wayfinding behavior and spatial knowledge acquisition: are they the same in virtual reality and in real-world environments? *Ann. Am. Assoc. Geogr.* 112 (1), 226–246.

Duran, I.E., 2016. The influence of familiarity and signage on wayfinding in academic libraries. The case of Bilkent University Library, Bilkent Universitesi (Turkey).

Feng, Y., et al., 2021. Using virtual reality to study pedestrian exit choice behaviour during evacuations. *Saf. Sci.* 137, 105158.

Feng, Y., et al., 2022a. Development and evaluation of a VR research tool to study wayfinding behaviour in a multi-story building. *Saf. Sci.* 147, 105573.

Feng, Y., et al., 2022b. Wayfinding behaviour in a multi-level building: a comparative study of HMD VR and Desktop VR. *Adv. Eng. Inf.* 51, 101475.

Fu, M., et al., 2023. How individuals sense environments during indoor emergency wayfinding: an eye-tracking investigation. *Journal of Building Engineering* 79, 107854.

Hassan, M., et al., 2025. Integration of Extended reality Technologies in Transportation Systems: a Bibliometric Analysis and Review of Emerging Trends, challenges, and Future Research. *Results Eng.* 105334.

He, Q., et al., 2019. Acquisition and transfer of spatial knowledge during wayfinding. *J. Exp. Psychol. Learn. Mem. Cogn.* 45 (8), 1364.

Hejtmánek, L., et al., 2020. How much of what we learn in virtual reality transfers to real-world navigation? *Multisens. Res.* 33 (4–5), 479–503.

Hölscher, C., et al., 2009. Adaptivity of wayfinding strategies in a multi-building ensemble: the effects of spatial structure, task requirements, and metric information. *J. Environ. Psychol.* 29 (2), 208–219.

Hölscher, C., et al., 2006. Up the down staircase: Wayfinding strategies in multi-level buildings. *J. Environ. Psychol.* 26 (4), 284–299.

Hund, A.M., Nazarczuk, S.N., 2009. The effects of sense of direction and training experience on wayfinding efficiency. *J. Environ. Psychol.* 29 (1), 151–159.

Iftikhar, H., Abbas, F., 2026. Conceptual models, influencing factors, and technological advancements in pedestrian wayfinding: a systematic literature review. *Facilities* 1–20.

Iftikhar, H., et al., 2021. A cross-cultural investigation of design and visual preference of signage information from Hong Kong and Pakistan. *The Journal of Navigation* 74 (2), 360–378.

Jamieson, S., 2004. Likert scales: how to (ab) use them? *Med. Educ.* 38 (12), 1217–1218.

Jiang, X., et al., 2015. Analyses of pedestrian behavior on mid-block unsignalized crosswalk comparing chinese and German cases. *Adv. Mech. Eng.* 7 (11), 1687814015610468.

Jiayu, Z., et al. (2025). "Influence of Different Multimodal Alarm Systems on Evacuation Behavior a Cross-Cultural Immersive Virtual Reality-Based Study." Available at SSRN 5148529.

Kaminka, G.A., Fridman, N., 2018. Simulating urban pedestrian crowds of different cultures. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9 (3), 1–27.

Kaplan, A.D., et al., 2021. The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: a meta-analysis. *Hum. Factors* 63 (4), 706–726.

Kattenbeck, M., et al., 2025. Predicting spatial familiarity by exploiting head and eye movements during pedestrian navigation in the real world. *Sci. Rep.* 15 (1), 7970.

Kennedy, R.S., et al., 1993. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3 (3), 203–220.

Kim, J.Y., et al., 2021. The influence of users' spatial familiarity on their emotional perception of space and wayfinding movement patterns. *Sensors* 21 (8), 2583.

Kim, K., Bock, O., 2021. Acquisition of landmark, route, and survey knowledge in a wayfinding task: in stages or in parallel? *Psychol. Res.* 85 (5), 2098–2106.

Li, R. (2012). "Toward a Holistic Framework of Indoor Wayfinding Behaviors: The Impacts of Architectural Intelligibility and Individual Differences."

Li, R., Klippel, A., 2016. Wayfinding behaviors in complex buildings: the impact of environmental legibility and familiarity. *Environ. Behav.* 48 (3), 482–510.

Lin, J., et al., 2019. Assessing the influence of repeated exposures and mental stress on human wayfinding performance in indoor environments using virtual reality technology. *Adv. Eng. Inf.* 39, 53–61.

Lin, Y. and A. Suh (2021). "The role of spatial ability in learning with virtual reality: a literature review."

Lu, Y., Bozovic-Stamenovic, R., 2009. Cultural perspective of wayfinding behavior: Exploring the socio-spatial variable in three chinese hospital case studies. *International Journal of Architectural Research* 3 (2), 22–34.

Meilinger, T., et al., 2006. How much information do you need? Schematic maps in wayfinding and self localisation. *International Conference on Spatial Cognition*.

Montana, J.I., et al., 2019. Neurorehabilitation of spatial memory using virtual environments: a systematic review. *J. Clin. Med.* 8 (10), 1516.

Muffato, V. (2016). "Spatial learning from maps and navigation: the role of age and visuo-spatial factors."

Muffato, V., Meneghetti, C., 2020. Knowledge of familiar environments: Assessing modalities and individual visuo-spatial factors. *J. Environ. Psychol.* 67, 101387.

Münzer, S., Stahl, C., 2011. Learning routes from visualizations for indoor wayfinding: Presentation modes and individual differences. *Spatial Cognition & Computation* 11 (4), 281–312.

Münzer, S., et al., 2012. Navigation assistance: a trade-off between wayfinding support and configural learning support. *J. Exp. Psychol. Appl.* 18 (1), 18.

O'Neill, M.J., 1992. Effects of familiarity and plan complexity on wayfinding in simulated buildings. *J. Environ. Psychol.* 12 (4), 319–327.

Park, G., 2013. "The Influence of Cultural Background on Wayfinding Cues. In Unfamiliar Buildings."

Raubal, M., 2001. Human wayfinding in unfamiliar buildings: a simulation with a cognizing agent. *Cogn. Process.* 2 (3), 363–388.

Reggente, N., et al., 2020. The method of loci in virtual reality: explicit binding of objects to spatial contexts enhances subsequent memory recall. *J. Cognit. Enhancement* 4 (1), 12–30.

Riva, G., Serino, S., 2020. Virtual reality in the assessment, understanding and treatment of mental health disorders. *MDPI* 9, 3434.

Sacco, K., et al., 2022. A virtual navigation training promotes the remapping of space in allocentric coordinates: evidence from behavioral and neuroimaging data. *Front. Hum. Neurosci.* 16, 693968.

Sauzéon, H., et al., 2016. Age and active navigation effects on episodic memory: a virtual reality study. *Br. J. Psychol.* 107 (1), 72–94.

Schmitz, S., 1997. Gender-related strategies in environmental development: Effects of anxiety on wayfinding in and representation of a three-dimensional maze. *J. Environ. Psychol.* 17 (3), 215–228.

Serrano, S.M., et al., 2024. Cross-cultural analysis of pedestrian group behaviour influence on crossing decisions in interactions with autonomous vehicles. 2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC), IEEE.

Sprenger, J., et al., 2023. Cross-cultural behavior analysis of street-crossing pedestrians in japan and germany. 2023 IEEE Intelligent Vehicles Symposium (IV), IEEE.

Stites, M.C., et al., 2020. Where are we going and where have we been? Examining the effects of maps on spatial learning in an indoor guided navigation task. *Cognit. Res.: Princ. Implic.* 5 (1), 13.

Sueur, C., et al., 2013. Different risk thresholds in pedestrian road crossing behaviour: a comparison of French and Japanese approaches. *Accid. Anal. Prev.* 58, 59–63.

- Sun, R., et al., 2019. The effect of a virtual reality learning environment on learners' spatial ability. *Virtual Reality* 23 (4), 385–398.
- Thomas, K.G., et al., 2010. Stress-induced impairment of spatial navigation in females. *South African Journal of Psychology* 40 (1), 44–53.
- Tuker, C., Tong, T., 2026. Spatial Layout Learning in Architecture: Comparative Media Insights from Field Trips, VR, and Video. *Visual Communication Quarterly*(just-Accepted) 1–18.
- van Beek, A., et al., 2024. Comparison of pedestrian wayfinding behavior between a real and a virtual multi-story building—a validation study. *Transp. Res. Part C Emerging Technol.* 163, 104650.
- Van der Wal, C. N., et al. (2024). “The effect of national culture on evacuation response behaviour: a cross-cultural survey.”.
- Veeraswamy, A., et al., 2011. Wayfinding behavior within buildings-an international survey. *Fire Saf. Sci.* 10, 735–748.
- Verghote, A., et al., 2019. The effects of information format and spatial cognition on individual wayfinding performance. *Buildings* 9 (2), 29.
- Wallet, G., et al., 2011. Virtual/real transfer of spatial knowledge: Benefit from visual fidelity provided in a virtual environment and impact of active navigation. *Cyberpsychol. Behav. Soc. Netw.* 14 (7–8), 417–423.
- Wallet, G., et al., 2009. Transfer of spatial knowledge from a virtual environment to reality: Impact of route complexity and subject's strategy on the exploration mode. *Journal of Virtual Reality and Broadcasting* 6.
- Witmer, B.G., et al., 1996. Virtual spaces and real world places: transfer of route knowledge. *Int. J. Hum Comput Stud.* 45 (4), 413–428.
- Witmer, B.G., et al., 2005. The factor structure of the presence questionnaire. *Presence: Teleoperators & Virtual Environments* 14 (3), 298–312.
- Ye, Y., et al., 2020. Risks to pedestrians in traffic systems with unfamiliar driving rules: a virtual reality approach. *Accid. Anal. Prev.* 142, 105565.
- Yu, L., et al., 2023. Subway passengers' wayfinding behaviors when exposed to signage: an experimental study in virtual reality with eye-tracker. *Saf. Sci.* 162, 106096.
- Zhang, M., et al., 2021. Investigating the influence of route turning angle on compliance behaviors and evacuation performance in a virtual-reality-based experiment. *Adv. Eng. Inf.* 48, 101259.
- Zhu, R., et al., 2020. Influence of architectural visual access on emergency wayfinding: a cross-cultural study in China, United Kingdom and United States. *Fire Saf. J.* 113, 102963.