Journal of the Brazilian Computer Society, 2026, 32:1, doi: 10.5753/jbcs.2026.8242

Be Yourself Driving: User Factors in a Real-Vehicle XR Simulator

Gustavo Camargo Domingues [**Voxar Labs, Centro de Informática, UFPE** | ***gcd2@cin.ufpe.br***]
David Ohara de Souza Cardoso [**Voxar Labs, Centro de Informática, UFPE** | ***dosc@cin.ufpe.br***]
Carlos Gabriel Bezerra Pereira [**Voxar Labs, Centro de Informática, UFPE** | ***cgbp@cin.ufpe.br***]
Marina Holanda Kunst [**Universidade Federal de Pernambuco** | ***marina.kunst@ufpe.br***]
Willams de Lima Costa [**Voxar Labs, Centro de Informática, UFPE** | ***wlc2@cin.ufpe.br***]
Veronica Teichrieb [**Voxar Labs, Centro de Informática, UFPE** | ***vt@cin.ufpe.br***]
Fátima de Lourdes dos Santos Nunes Marques [**Universidade de São Paulo**| ***fatima.nunes@usp.br***]
João Marcelo Teixeira [**Universidade de São Paulo, Universidade Federal de Pernambuco**| ***jmxnt@cin.ufpe.br***]

Center of Informatics, Universidade Federal de Pernambuco, Av. Jornalista Aníbal Fernandes, s/n, Cidade Universitária, Recife, PE, 50740-560, Brazil.

Received: 17 May 2026 • **Accepted:** 14 June 2026 • **Published:** 23 July 2026

Abstract Real-vehicle Extended Reality (XR) simulators are attractive because they preserve the physical steering wheel, pedals, and cabin while enabling controlled virtual scenarios. However, automotive XR studies often emphasize condition-level comparisons, such as Mixed Reality (MR) versus Virtual Reality (VR), while giving less attention to participant factors that may shape objective performance. This paper investigates how gender, driver's license status, and practical driving experience relate to performance in a physically grounded XR driving task implemented inside a stationary BYD Dolphin Plus. Forty participants, balanced by gender and license status, completed the same collectible-based driving task in both VR and MR, yielding 80 logged runs. Objective measures included batteries collected, first-lap duration, and percentage of time spent inside the track. Linear mixed-effects models showed that driver's license status was a key predictor of performance: licensed participants collected more batteries and remained inside the track for longer. Gender-related effects also appeared for batteries collected and track adherence, but subgroup and licensed-only analyzes indicated that these effects should not be interpreted as intrinsic gender differences; instead, they were intertwined with license status, practical exposure, and the higher variability of the unlicensed subgroup. Visualization condition mainly affected first-lap duration, with shorter laps in VR, while batteries collected and track adherence remained comparatively stable across conditions. Exploratory analyses further suggested that, among licensed drivers, recent practical driving activity was more informative than gender for explaining performance differences. We discuss the implications of these findings for automotive XR evaluation, participant profiling, and the design of real-vehicle XR simulators.

Keywords: Real-vehicle XR simulator, automotive extended reality, user factors, driving experience, cybersickness.

1 Introduction

Extended Reality (XR) is increasingly used in automotive research for training, simulation, prototyping, and in-vehicle interaction studies [Kemeny, 2014; Schroeter and Gerber, 2018; Silvera *et al*., 2022; Hassan *et al*., 2025]. A central promise of XR in this domain is experimental control: researchers can manipulate road layouts, obstacles, feedback, and interface conditions without the cost, safety constraints, or logistical complexity of on-road experimentation [Morra *et al*., 2019; Mayhew *et al*., 2011]. At the same time, the usefulness of an automotive XR setup depends heavily on how convincingly it preserves embodied interaction with the vehicle. Steering, braking, and speed regulation are not purely visual tasks; they depend on tightly coupled sensorimotor feedback from the body, the controls, and the surrounding cabin [Hock *et al*., 2017; Colley *et al*., 2022; Hock *et al*., 2022].

Recent systems have therefore moved beyond desktop or wheel-and-pedal simulators toward more physically grounded configurations. Examples include in-car passenger XR platforms [McGill *et al*., 2022], MR on-road concepts [Ghiurău *et al*., 2020; Goedicke *et al*., 2021], and affordable research-oriented VR driving environments [Schroeter and Gerber, 2018; Silvera *et al*., 2022]. This shift is important because it reframes simulation fidelity as more than graphics or motion rendering. In automotive XR, fidelity also depends on whether a system sustains credible interaction between the real driver, the physical controls, and the virtual scene [Himmels *et al*., 2022; de Winkel *et al*., 2022].

In parallel, the driving literature has long shown that individual differences matter. Gender has been discussed in relation to presence, virtual-driver perception, and simulator response [Darty *et al*., 2014]. Driver expertise and recency of driving practice also influence behavior, anticipation, and performance in simulated environments [Mayhew *et al*., 2011; Stahl *et al*., 2016]. Yet these user factors remain underexplored in real-vehicle XR simulators, where motor coordination, proprioception, and visual-physical alignment can interact in complex ways.

This paper analyzes a real-vehicle XR driving setup from a user-factor perspective. Instead of asking only whether VR or MR is globally better, we ask how participant background relates to objective performance when the physical car is preserved as the interaction substrate. Our experiment used a stationary BYD Dolphin Plus equipped with OBD-II telemetry and a Meta Quest 3 headset. Participants completed the same driving-and-collection task in both a fully virtual condition and a mixed-reality condition that preserved key real-car references while virtualizing the outside world. The dataset comprises 40 participants balanced by gender and license status, and includes both objective logs and post-condition questionnaires.

The paper makes three contributions:

1. It provides a focused analysis of how gender and driver's license status relate to objective performance in a real-vehicle XR simulator.
2. It examines self-reported practical experience and driving recency as exploratory factors beyond the binary licensed/unlicensed distinction, while using cybersickness and open-ended comments as complementary evidence.
3. It translates the resulting patterns into design implications for embodied automotive XR systems, especially regarding calibration, control sensitivity, and participant profiling.

The remainder of this paper is organized as follows. Section 2 reviews related work on XR in automotive research, simulation fidelity, embodiment, and user factors in driving contexts. Section 3 presents the real-vehicle XR simulator, including the vehicle setup, telemetry pipeline, visualization conditions, and calibration procedure. Section 4 describes the study design, participants, task, measures, and statistical analysis strategy. Section 5 reports the objective performance, practical-experience, open-ended comment, and cybersickness results. Section 6 discusses the implications of the findings for real-vehicle XR evaluation and participant profiling. Finally, Section 7 concludes the paper, summarizes its limitations, and outlines directions for future work.

2 Related Work

2.1 XR in automotive research

Driving simulation has evolved from desktop-style displays and simplified controls toward immersive XR systems that support richer interaction and more ecologically grounded studies [Kemeny, 2014; Schroeter and Gerber, 2018]. In automotive HCI, XR is now used both to reproduce traffic scenes and to study how drivers perceive information, interact with controls, and respond to novel interfaces [Morra *et al.*, 2019; Silvera *et al.*, 2022].

XR has also become relevant across several automotive activities, including vehicle design, testing and simulation, training, education, marketing, and sales [Postelnicu and Boboc, 2024]. A growing line of work has specifically explored XR in real vehicles or vehicle-like settings. Hock *et al.* [2017] showed that in-car VR can support compelling entertainment experiences when physical motion cues are preserved. McGill *et al.* [2022] introduced an open platform for any-car XR experiences, demonstrating that real vehicle movement can be synchronized with virtual content at low cost. Mixed-reality concepts such as ARCAR [Ghiurău *et al.*, 2020] and XR-OOM [Goedicke *et al.*, 2021] further suggest that integrating real cars with virtual elements can support novel design spaces for experimentation, prototyping, and driver assistance. These systems are especially relevant because XR combines technological immersion with users' subjective sense of presence, which can shape how convincingly participants experience and respond to simulated driving scenarios [Postelnicu and Boboc, 2024; Lønne *et al.*, 2023].

2.2 Fidelity, embodiment, and automotive XR

In VR and MR simulations, users interact with both physical and virtual elements, making calibration and spatial alignment central to the quality of the experience. When virtual prototypes provide visual, tactile, and auditory feedback, including graphical details such as colors, textures, and animated components, they can support a more immersive experience [Winkler *et al.*, 2022]. In driving simulation, however, fidelity is multidimensional. It involves not only graphics and field of view, but also how well the simulator reproduces speed perception, vehicle dynamics, and the embodied link between physical action and perceived response [Bella, 2008; Törnros, 1998; Hussain *et al.*, 2019]. Prior work has shown that greater motion realism and richer bodily feedback can improve presence and behavioral validity [Colley *et al.*, 2022; Hock *et al.*, 2022]. Conversely, miscalibrated or poorly coupled setups may degrade interaction quality and increase cybersickness [de Winkel *et al.*, 2022].

This issue is especially important in real-vehicle XR. If the steering wheel, pedals, body posture, and cabin are real, then misalignment between physical and virtual representations may influence perceived realism, steering stability, and control confidence even when the underlying input signal is correct [Goedicke *et al.*, 2021; Himmels *et al.*, 2022]. Real-vehicle XR simulators therefore offer an important testbed for studying whether performance differences are better explained by visualization mode, system calibration, or participant background and skill.

2.3 Gender, license status, and driving experience

The role of gender in simulated driving should be interpreted cautiously. Prior work has reported differences in presence and virtual-driver perception, but the literature does not support simplistic claims that one gender is inherently better at driving in general [Darty *et al.*, 2014]. Instead, observed differences often depend on the task, the metric, and the simulator configuration.

By contrast, differences between novice and experienced drivers are theoretically better grounded. Experience affects visual exploration, anticipation, risk perception, and adaptation to new driving tasks [Stahl *et al.*, 2016; Mayhew *et al.*, 2011]. Aldoski [2025] notes that VR-based training can reduce control-taking reaction times and improve procedural accuracy. In XR, these differences may become even more consequential because users must simultaneously manage driving actions and adapt to a mediated visual environment.

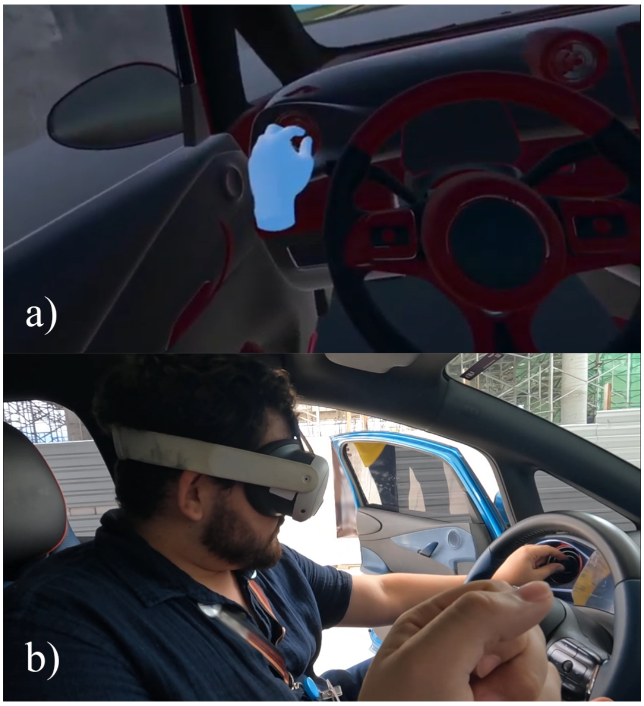

Figure 1. Real-vehicle XR simulator. The virtual cockpit view (a) was spatially matched to the physical vehicle setup (b), allowing participants to interact with the real steering wheel and pedals while receiving passive haptic feedback from the actual car interior.

Driver's license status is therefore a useful, but coarse, proxy for prior vehicle-control experience. Practical exposure and recency of real driving may explain additional variance even among licensed drivers. This motivates analyzing license status as a primary participant factor while treating self-reported practical experience and driving recency as exploratory secondary dimensions.

3 Real-Vehicle XR Simulator

The experimental platform was built around a stationary BYD Dolphin Plus. Participants sat inside the physical vehicle, used the original steering wheel and pedals, and viewed the task through a Meta Quest 3 headset. The vehicle remained stationary throughout the experiment, but the engine and air conditioning were kept on to preserve a realistic cabin state and participant comfort. In this setup, the physical cockpit served as the interaction substrate, while the headset controlled which visual elements of the vehicle and driving environment were presented to the participant. Figure 1 shows the correspondence between the virtual cockpit view and the physical in-car setup.

Vehicle telemetry was acquired through the OBD-II port using an ELM327 Bluetooth adapter connected to the vehicle's CAN network. Because standard OBD-II PID access did not expose the required control signals, accelerator position, brake position, and steering-wheel angle were read using Unified Diagnostic Services over ISO 15765-4. The acquired signals were transmitted to the Meta Quest 3 via Bluetooth and accessed inside the Unity application through an Android/Java bridge. These signals drove the virtual vehicle during each run and allowed the objective task logs to be associated with the participant's physical control inputs.

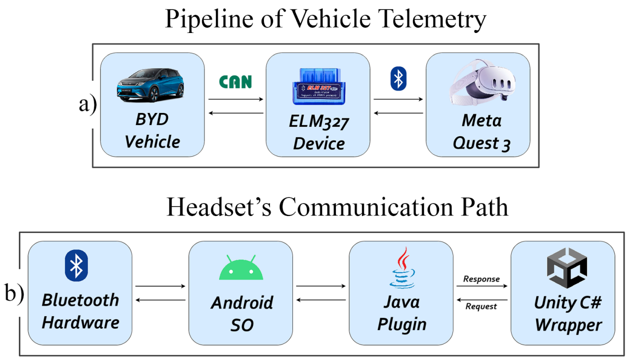

Figure 2. Vehicle–OBD-II–Quest 3 communication workflow. The ELM327 device established communication between the vehicle and the headset (a). The internal processing path converted vehicle data into application variables through Bluetooth communication, Android/Java integration, and a Unity wrapper (b).

The application rendered the same task environment in two visualization conditions. In the VR condition, the surrounding scene was fully virtual, although participants still interacted with the real steering wheel and pedals. In the MR condition, key visual references from the real vehicle context remained visible while the outside driving environment and task elements were virtual. Both conditions used the same track, the same collectibles, and the same task logic.

Before each run, a researcher manually aligned the virtual steering wheel with the real one. The steering wheel was used as the primary calibration anchor because it was the main control surface for the task and the most important reference for hand placement, steering input, and perceived spatial correspondence. In the MR condition, the surrounding car windows were also used as additional visual references to stabilize scene placement relative to the vehicle interior. Because steering in the virtual scene was driven by the real steering-wheel angle acquired from the vehicle, calibration primarily affected perceived visual-physical correspondence rather than the underlying steering signal.

The logging pipeline synchronized scene-level task measures with vehicle telemetry using a shared session timestamp. Scene-level data were logged in the application at 10 Hz and included elapsed time, collected batteries, first-lap duration, and time spent inside and outside the track. This enabled offline computation of the objective performance measures analyzed in this paper, as shown in Figure 2.

4 Method

This section describes the study design, participant sample, task, measures, and analysis strategy used to examine how user factors relate to objective performance in the real-vehicle XR simulator. Although each participant completed both visualization conditions, the present analysis treats condition as a repeated-measures factor so that performance differences associated with gender, driver's license status, practical driving experience, and driving recency can be evaluated while accounting for repeated runs from the same participant.

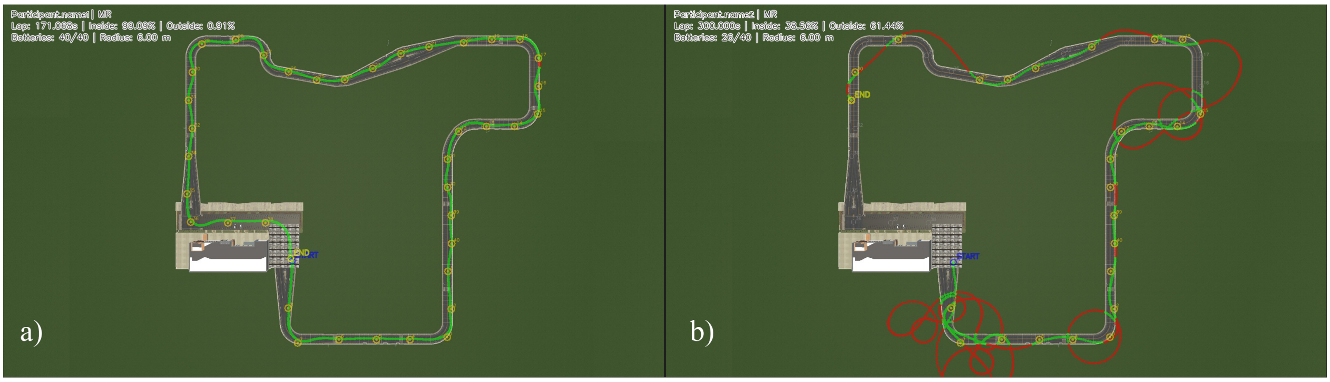

Figure 3. Example per-run trajectory maps generated from the objective logs. Each map summarizes the visualization condition, first lap completion time, proportion of time spent inside and outside the drivable area, battery collection, and the recorded trajectory. The examples contrast a high-adherence run (a) with a run showing unstable curve handling and frequent departures from the track (b), both in the MR condition.

4.1 Study design

We conducted a within-subject experiment in which each participant completed the same driving task in two visualization conditions: VR and MR. Both conditions used the same virtual environment, track geometry, collectible placement, obstacle placement, and task duration. The order of conditions was counterbalanced by alternating assignment across participants, so that successive participants completed the conditions in opposite orders. This design allowed visualization condition to be analyzed as a repeated-measures factor while keeping participant-level factors as the main focus of the paper.

In the VR condition, all visual content was rendered as a fully virtual environment. Participants' hands were highlighted in blue to facilitate spatial awareness and indicate their position within the virtual scene. In the MR condition, the real vehicle interior was visible to the participant, whereas the exterior environment remained virtual. The blue hand-highlight visualization was preserved across both conditions to maintain consistent hand-position feedback.

4.2 Participants

The dataset comprises 40 participants from the Informatics Center of the Federal University of Pernambuco. The sample was balanced by gender and driver's license status: 20 women and 20 men; 20 licensed and 20 unlicensed participants (Table 1). Crossing gender and license status yielded four equally sized groups of 10 participants each. Mean age was 24.6 years (SD = 5.8). Self-reported familiarity with VR was moderate on average (M = 3.15, SD = 1.53 on a 1–5 scale), while familiarity with vehicle simulators was lower (M = 2.02, SD = 1.59).

For exploratory analyzes, self-reported driving history was transformed into three practical-experience groups: no practical experience (20 participants), up to 2 years of experience (10 participants), and more than 2 years of experience (10 participants). This grouping reflects the natural structure of the questionnaire responses and should be read as an exploratory categorization rather than as a predefined competence threshold. The characterization questionnaire also collected age and education level. We also extracted the reported time since the participant last drove, which was later used as a recency proxy in the licensed-only analyses.

This study was approved by the Research Ethics Committee of the Federal University of Pernambuco under number 95417326.4.0000.5208.

4.3 Task and procedure

The task abstracted driving into a controlled XR challenge. Participants navigated a virtual track while trying to collect as many batteries as possible and remain inside the road boundaries. The task required continuous steering, speed modulation, and obstacle avoidance, but intentionally avoided traffic complexity so that the analysis could focus on embodied vehicle control in the XR setup.

At the beginning of the session, participants gave their informed consent and completed the characterization questionnaire. Then, they were seated in the experimental setup and positioned relative to the vehicle and steering wheel. Before each condition, the XR system and the steering-wheel alignment were calibrated for the current run. Participants received instructions on the controls, completed a one-minute familiarization phase (present only on the first visualization moment), and then performed the five-minute task while telemetry and video were recorded. After each run, participants completed a post-condition questionnaire about their experience in that condition. The second visualization condition was then presented, and the same calibration, task, and post-condition assessment steps were repeated.

4.4 Measures

The analyzes used two data sources. Objective performance measures were extracted from the simulator logs, with one row per participant and visualization condition. Participant-background variables were extracted from the characterization questionnaire completed before the XR conditions.

We analyzed three primary objective performance outcomes from the simulator logs:

1. **Batteries collected**: total number of collectibles obtained during the run.
2. **First-lap duration**: time required to complete the first full lap of the course.
3. **Inside-track percentage**: percentage of sampled time spent inside the drivable area.

Table 1. Participant profile. Asterisks indicate ratings on a 1–5 scale.

Characteristic	Value
Participants	40
Women / Men	20 / 20
Licensed / Unlicensed	20 / 20
Age, mean (SD)	24.6 (5.8) years
VR familiarity*, mean (SD)	3.15 (1.53)
Simulator familiarity*, mean (SD)	2.02 (1.59)
Practical experience groups	20 none / 10 up to 2 years / 10 >2 years

The outside-track percentage was also logged, but because it is mathematically redundant with inside-track percentage, we use it only for interpretation and descriptive reporting. Figure 3 illustrates the type of trajectory-level information extracted from the objective logs.

The characterization questionnaire provided participant-background variables, including gender, driver's license status, self-reported practical driving experience, and time since last driving activity. Gender and license status were used in the primary objective-log analyses. Practical driving experience and driving recency were used only in exploratory participant-background analyses.

The post-condition questionnaire also included subjective ratings and an optional free-text comment field. In the present paper, cybersickness ratings and open-ended comments are used as complementary evidence rather than as primary outcomes. Cybersickness symptoms included dizziness or vertigo, visual fatigue, eye pain, headache, nausea, headset discomfort, and discomfort while driving, each rated on a 1–7 scale from low to high intensity. Open-ended comments were reviewed to identify recurring themes that helped contextualize the quantitative results.

4.5 Analysis strategy

The primary objective-log analysis used linear mixed-effects models to account for the repeated-measures structure of the experiment, since each participant completed both visualization conditions. For each objective metric, we fitted a model with participant as a random intercept and condition, gender, license status, and their interactions as fixed effects:

$$\text{metric} \sim \text{condition} * \text{gender} * \text{license} + (1|\text{participant}).$$

This model was run separately for batteries collected, first-lap duration, and inside-track percentage. Condition was treated as a within-participant factor, while gender and license status were treated as participant-level factors. Unless otherwise stated, MR, women, and unlicensed participants were used as the reference levels. The complete fixed-effect structure was inspected for each outcome so that condition effects, participant-background effects, and their interactions could be interpreted consistently.

As a complementary omnibus analysis, we also ran mixed ANOVAs with visualization condition as the within-participant factor and user profile as a between-participant factor. User profile was defined by the four gender $\times$ license subgroups: women without a license, women with a license, men without a license, and men with a license. These analyses were used to support the interpretation of subgroup patterns and to guide Bonferroni-corrected post-hoc comparisons. The mixed-effects models remained the primary analysis because they modeled repeated observations at the participant level and allowed gender and license status to be evaluated as separate factors.

Participant-background analysis involving practical driving experience and driving recency were treated as exploratory because these variables came from the characterization questionnaire and overlapped structurally with license status. Participants without a driver's license also belonged to the no-practical-experience group, so experience could not be interpreted independently from license status in the full sample. We therefore handled these variables in two steps. First, we compared the three practical-experience groups descriptively and with nonparametric tests. Second, among licensed participants only, we fitted exploratory regressions with standard errors clustered by participant:

$$\text{metric} \sim \text{condition} + \text{gender} + \text{driving\_experience\_months} + \text{months\_since\_last\_driving}.$$

This licensed-only analysis asked whether accumulated driving experience and driving recency explained additional variation once all participants in the subset already held a driver's license. Because this analysis used questionnaire-derived variables and a smaller subset of participants, it was interpreted as exploratory rather than as a primary inferential result.

Complementary learning analyses compared performance between the first and second experimental runs to contextualize adaptation across the session. Cybersickness ratings were summarized descriptively by participant subgroup and run order, and open-ended comments were used as interpretive support rather than as confirmatory evidence. All analyzes were conducted in Python via Google Colab. Where pairwise follow-up comparisons were used, p-values were adjusted using Bonferroni correction.

5 Results

We report the results in four steps. First, we provide descriptive summaries for the four gender $\times$ license subgroups. In the mixed-effects models reported later, gender and license status are treated as separate predictors. When we refer to the four-level *user profile* or *gender $\times$ license subgroup*, we mean the combined groups formed by crossing gender and license status: women without a license, women with a license, men without a license, and men with a license. This distinction is important because the primary analyses evaluate gender and license status separately, whereas the complementary omnibus and post-hoc analyses use the four subgroup profiles to interpret group-level patterns.

5.1 Descriptive overview

Table 2 summarizes the mean objective performance for the four gender $\times$ license subgroups under both visualization conditions. These descriptive results show two initial patterns. First, licensed participants generally collected more batteries and remained inside the track for longer than unlicensed participants. Second, the lowest mean performance appeared in the unlicensed women subgroup, especially for batteries collected and inside-track percentage. By contrast, the two licensed subgroups showed more similar performance to each other. These descriptive patterns are used only as an initial overview; the inferential analyses in the following subsections test whether these differences remain reliable when condition and repeated measures are modeled explicitly.

5.2 Mixed-effects models

Table 3 reports the complete fixed-effect structure from the mixed-effects models. In these models, gender and license status were entered as separate predictors, with participant modeled as a random intercept. MR, women, and unlicensed participants were used as reference levels. Therefore, the coefficients for men and licensed participants should be interpreted relative to the reference subgroup, and interaction terms indicate whether these effects changed across condition, gender, or license status.

For batteries collected, the coefficients for men ($\beta = 10.10$, $p < .001$) and licensed participants ($\beta = 11.70$, $p < .001$) were significant in the reference condition, while visualization condition and all interaction terms were not statistically significant. For inside-track percentage, the same general pattern emerged: men ($\beta = 15.26$, $p = .004$) and licensed participants ($\beta = 19.03$, $p < .001$) showed higher values relative to the reference subgroup, whereas condition did not reach significance. For first-lap duration, visualization condition was the only significant predictor, with shorter first laps in VR ($\beta = -32.51$ s, $p = .020$). No interaction term reached the conventional significance threshold in any model.

These models indicate that participant background was more informative than visualization condition for the two main performance-quality metrics: battery collection (Figure 4) and track adherence (Figure 5). As a complementary omnibus analysis, the mixed ANOVA used the four gender $\times$ license subgroups as a between-participant user-profile factor and visualization condition as the within-participant factor. This analysis showed a significant main effect of user profile ($p = 0.001$), while visualization condition alone did not show a significant effect on track adherence ($p = 0.995$). Although the descriptive means suggested a possible condition-dependent pattern, with women showing slightly higher values in VR and men showing slightly higher track-adherence values in MR, the user-profile $\times$ condition interaction did not reach statistical significance ($p = 0.276$). This pattern should therefore be interpreted as descriptive rather than confirmatory.

Bonferroni-corrected post-hoc comparisons further clarified the subgroup structure. The largest differences occurred between licensed and unlicensed participants rather than between women and men with the same license status. In the VR condition, licensed women showed higher track adherence than unlicensed men ($p = 0.018$, Hedges' $g = 1.604$). Conversely, licensed women and licensed men did not differ significantly in either condition ($p = 1.000$). These results suggest that the gender-related effects in the mixed-effects models should not be interpreted as evidence of an intrinsic gender difference in driving ability. Instead, the pattern appears to reflect the overlap between gender, license status, and practical exposure in this sample, especially the lower and more variable performance of the unlicensed women subgroup.

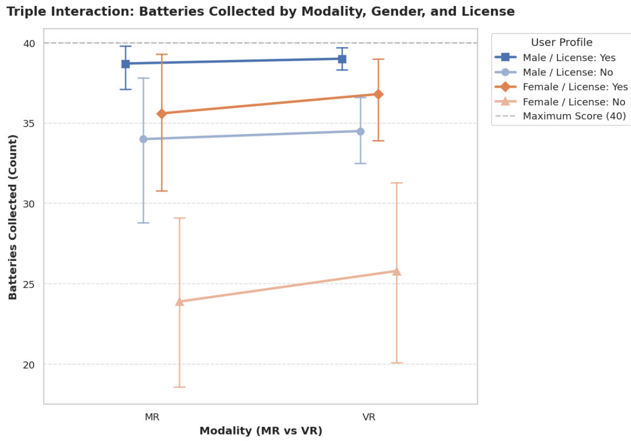

Figure 4. Battery collection by visualization condition, gender, and driver's license status.

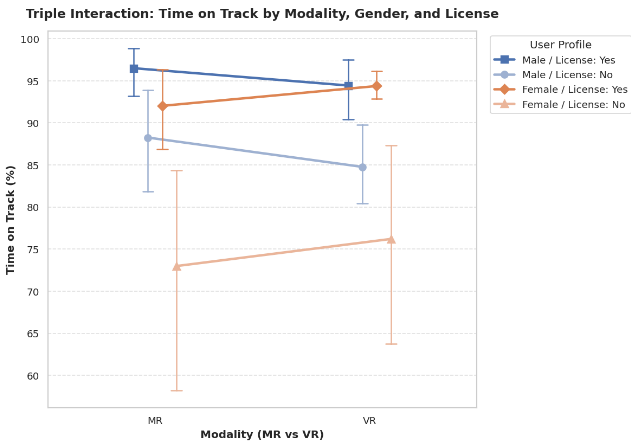

Figure 5. Inside-track percentage by visualization condition, gender, and driver's license status.

To check whether gender remained relevant once license status was held constant, we ran licensed-only regressions. In this subset, gender was not a significant predictor of batteries collected ($p = .117$) or inside-track percentage ($p = .321$). In other words, the gender-related pattern visible in Table 3 did not remain statistically reliable when the analysis was restricted to licensed drivers. This reinforces the interpretation that gender, license status, and practical exposure were

Table 2. Mean objective performance by condition, gender, and license status.

Gender	License	Batteries collected		Inside-track (%)		First-lap duration (s)	
		MR	VR	MR	VR	MR	VR
Women	No	23.9	25.8	72.99	76.21	200.19	167.68
Women	Yes	35.6	36.8	92.02	94.37	196.99	180.83
Men	No	34.0	34.5	88.25	84.77	181.47	173.88
Men	Yes	38.7	39.0	96.49	94.43	160.80	157.93

Means are computed from the objective logs. Because each participant completed both conditions, each cell contains 10 runs.

Table 3. Complete mixed-effects results with participant random intercept. MR, women, and unlicensed participants are the reference levels.

Outcome	Effect	β	SE	p
Batteries collected	VR condition	1.90	1.50	.206
	Men	10.10	2.74	<.001
	Licensed	11.70	2.74	<.001
	VR × Men	-1.40	2.12	.510
	VR × Licensed	-0.70	2.12	.742
	Men × Licensed	-7.00	3.87	.071
	VR × Men × Licensed	0.50	3.00	.868
Inside-track percentage	VR condition	3.22	2.69	.231
	Men	15.26	5.24	.004
	Licensed	19.03	5.24	<.001
	VR × Men	-6.70	3.80	.078
	VR × Licensed	-0.86	3.80	.820
	Men × Licensed	-10.79	7.40	.145
	VR × Men × Licensed	2.29	5.38	.670
First-lap duration	VR condition	-32.51	13.92	.020
	Men	-18.72	21.66	.388
	Licensed	-3.20	21.66	.883
	VR × Men	24.91	19.69	.206
	VR × Licensed	16.34	19.69	.407
	Men × Licensed	-17.48	30.63	.568
	VR × Men × Licensed	-11.62	27.85	.677

intertwined in this sample.

The dispersion analysis also showed that performance variability was concentrated among unlicensed participants. Licensed participants not only performed better on average, but also showed more consistent vehicle control: their standard deviation for inside-track percentage remained comparatively low across both visualization conditions, ranging from $SD = 2.87$ to $SD = 8.17$ (Figure 6). In contrast, unlicensed participants showed greater instability, particularly in the unlicensed women subgroup, which had the highest dispersion rates ($SD = 23.08$ in MR and $SD = 19.77$ in VR) and several outlying runs. This difference in variability suggests that practical driving exposure may have acted as a stabilizing factor in the real-vehicle XR simulator, reducing erratic performance regardless of visualization condition.

We also examined performance changes between the first and second experimental runs to assess short-term adaptation across the session. Track adherence showed a small descriptive improvement, but paired t-tests indicated that changes in inside-track percentage were not statistically significant for any subgroup ($p > 0.05$), with very small effect sizes (Cohen's $d < 0.25$), as shown in Figure 7.

First-lap duration showed clearer evidence of adaptation for some groups (Figure 8). Unlicensed participants significantly reduced their lap times in the second run ($p = 0.034$, $d = 0.496$), suggesting that they became faster as they adapted to the controls. Men also showed a significant reduction in lap duration ($p = 0.019$, $d = 0.418$). Licensed drivers, by contrast, showed little change in lap duration ($p = 0.856$, $d = 0.032$) and less than 1% improvement in track adherence. This pattern is consistent with a ceiling effect: participants with prior driving experience appeared able to transfer real-world control skills to the embodied XR setup with less within-session adaptation.

5.3 Practical experience and recency of driving

The practical-experience grouping also differentiated the sample. Across all participants, batteries collected differed significantly across the three experience groups (Kruskal–Wallis $H = 14.77$, $p < .001$), as did inside-track percentage ($H = 15.09$, $p < .001$). However, this separation mostly reflected the contrast between participants with no practical

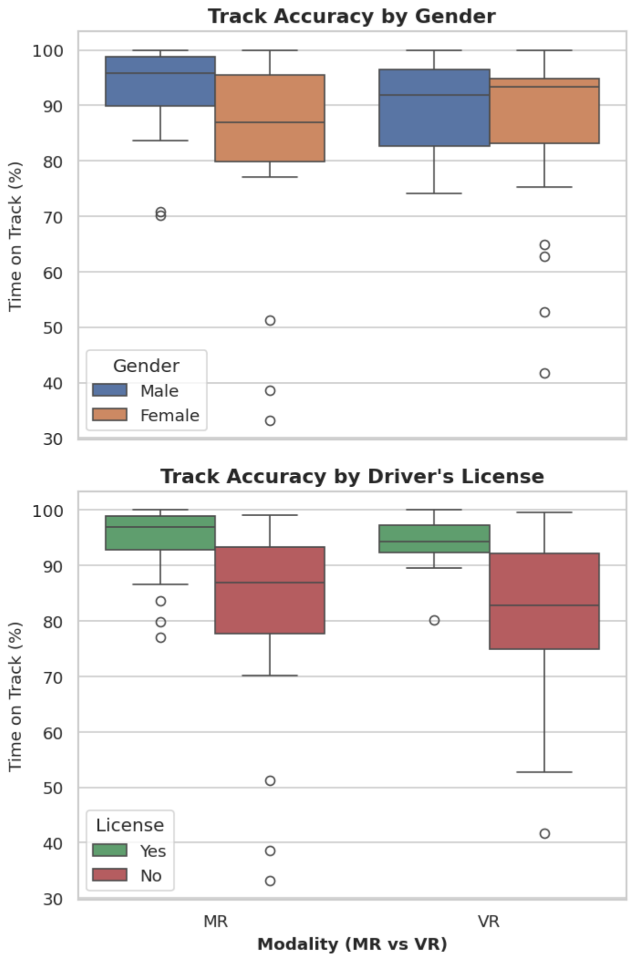

Figure 6. Inside-track percentage by visualization condition, gender, and driver's license status.

driving history and those with any driving exposure. As illustrated in Figure 9, the two experienced groups (up to 2 years and more than 2 years) were comparatively similar, with high median performance and low dispersion in both batteries collected and track adherence. Conversely, the group with no practical experience showed substantial variance and lower overall performance across both metrics.

Because practical experience overlapped structurally with license status, we also examined a licensed-only model including total driving history and driving recency. In that model, gender was again non-significant, while accumulated experience and recency showed meaningful relationships. More months of accumulated driving experience predicted slightly more batteries collected ($\beta = 0.010$, $p = .021$) and more time inside the track ($\beta = 0.026$, $p = .001$). By contrast, more months since the participant last drove predicted fewer batteries collected ($\beta = -0.086$, $p < .001$) and less time inside the track ($\beta = -0.080$, $p < .001$). Recency was therefore the stronger practical-experience indicator in this dataset.

Greater accumulated experience was also associated with longer first-lap duration among licensed participants ($\beta = 0.475$, $p < .001$), suggesting a more cautious or controlled driving style rather than a straightforward speed advantage. This finding should be interpreted cautiously, but it reinforces the idea that speed alone is not an adequate proxy for quality in this task.

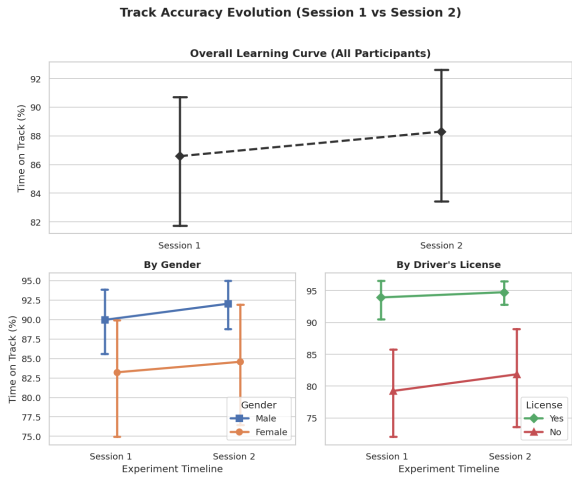

Figure 7. Inside-track percentage across the first and second experimental runs.

5.4 Open-ended comments

At the end of each post-condition questionnaire, participants could answer an optional open-ended questionnaire about their experience in the tested condition. These comments were used to contextualize the quantitative results, not as a formal qualitative coding study. Participants were invited to freely describe their real-vehicle XR driving experience, including what they liked or disliked, which features they thought should be improved, and how they felt while using the application.

We summarize the comments in two ways. First, we describe recurring themes across the VR and MR conditions (Subsection 5.4.1). Second, we discuss descriptive comment patterns related to gender, previous VR familiarity, and driving experience in the following subsections. These patterns should be interpreted as complementary observations rather than as statistical correlations.

5.4.1 Participants' feelings across simulator versions

The VR condition revealed a set of recurrent perceptions related to immersion, control, discomfort, realism, and adaptation. The main themes identified in the responses were:

- Thirteen participants described VR as immersive, interesting, fun, intriguing, or generally engaging, often emphasizing the novelty of the experience and the feeling of being inside the simulated environment.
- Thirteen participants reported control-related problems, including pedal sensitivity, abrupt deceleration after releasing the accelerator, limited steering responsiveness, and vehicle behavior that did not fully match their expectations of a real car.
- Seven participants highlighted real–virtual mismatches involving the external vehicle body, mirror position, hand placement, or broader cockpit desynchronization.
- Eight participants reported discomfort, especially nausea, dizziness, visual strain, or disorientation.
- Four participants described gradual adaptation, such as understanding the controls better, becoming more accustomed to the track, or perceiving the simulator as less imprecise over time.

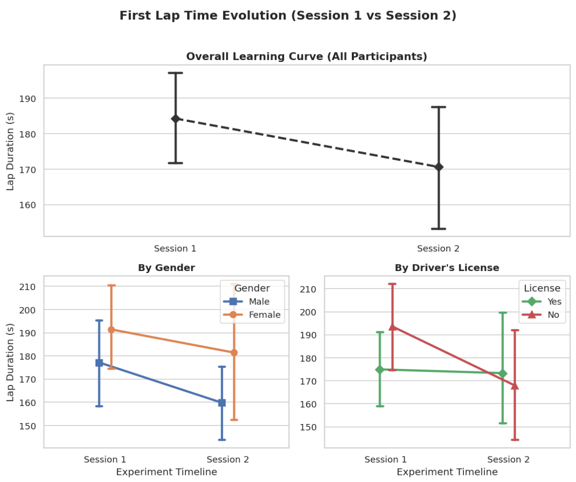

Figure 8. First-lap duration across the first and second experimental runs.

- Six participants mentioned the absence of bodily force or inertia during movement, as well as reduced visual realism.
- One participant framed VR as promising for driving training, including remote training possibilities, while another suggested that the introductory route should be expanded to improve familiarization before the main task.

In the MR condition, participants' comments also revealed recurring themes, but with a different emphasis. The visibility of the real vehicle appeared to influence how participants described realism, control, orientation, and overall usability. The main themes were:

- Fourteen participants described MR as enjoyable, interesting, convincing, fun, or immersive, often emphasizing that the condition was strong as an overall experience.
- Six participants felt that seeing the real vehicle improved orientation, movement perception, or control during the task.
- Nine participants mentioned control-related issues, including pedal sensitivity, abrupt braking, unrealistic slowing after releasing the accelerator, and steering behavior that still felt strange or inconsistent.
- Three participants reported difficulty with curves or with the task, often relating this difficulty to limited driving skill.
- Six participants reported discomfort, including dizziness, nausea, unease, or perceptual mismatches.
- Four participants mentioned batteries, checkpoints, or the collection task either as engaging elements or as elements that diverted attention away from driving.
- Two participants described MR-specific confusion or reduced immersion.
- Two participants mentioned adaptation over time.
- One participant suggested that MR could be useful for introductory driving lessons, while another described MR as reducing fear of taking the wheel compared with real-world driving.
- One participant highlighted the absence of sound feedback during collisions.

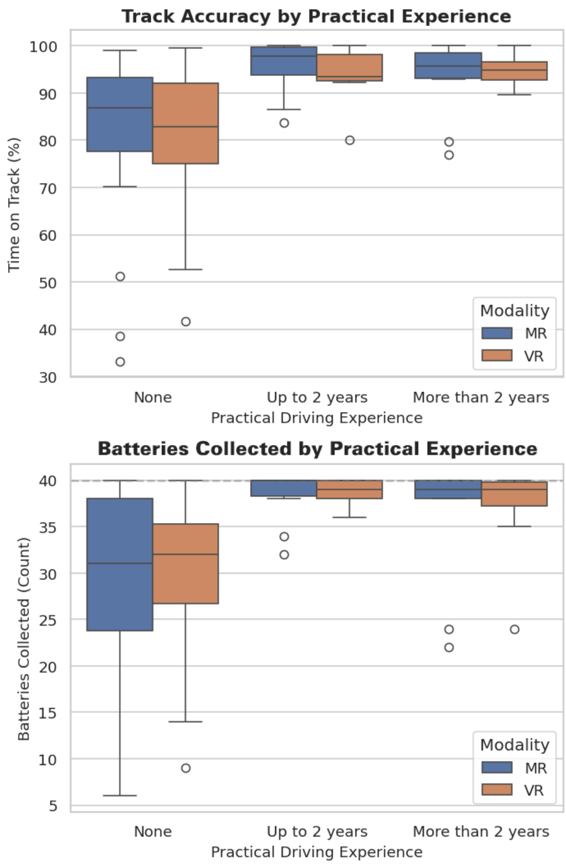

Figure 9. Exploratory effects of practical driving experience on inside-track percentage and batteries collected.

5.4.2 Gender-related comment patterns

Gender-related comments suggest different emphases in how men and women interpreted the simulator experience, although both groups discussed immersion, control, calibration, and discomfort. In the VR condition, men more often evaluated the experience through technical fidelity and vehicle realism. Their comments frequently addressed whether the virtual vehicle behaved like a real car, whether acceleration and braking were physically plausible, and whether the absence of inertia, gravity, or sound feedback limited realism. Some male participants also described the simulation as intriguing or enjoyable despite these limitations, indicating that technical criticism did not necessarily correspond to a negative overall evaluation. In the MR condition, male participants more often described the visibility of the real car as beneficial for orientation and control. They mentioned improved distance perception, greater confidence in vehicle position, and a stronger sense that the real and virtual elements formed a coherent driving environment. However, they also continued to report pedal sensitivity, abrupt braking, speed-perception problems, and alignment issues involving mirrors, the steering wheel, or the vehicle body.

Women's comments more often connected the experience with control difficulty, physical discomfort, and adaptation. In the VR condition, several women reported difficulty with curves, accelerator sensitivity, slow or imprecise steering response, abrupt deceleration, or uncertainty in judging how much pressure was being applied to the pedals. Some of these comments were explicitly associated with limited driving experience, while others referred to the simulator's behavior itself. Women also produced more comments related to nausea, dizziness, visual discomfort, and perceptual mismatch, which is consistent with the cybersickness analysis showing higher increases in dizziness, nausea, and visual fatigue among women from the first to the second round. In the MR condition, women often recognized the benefit of seeing the real vehicle, especially for physical orientation and immersion, but they still reported difficulty with curves, unusual acceleration behavior, and occasional disorientation. Therefore, these gender-related comment patterns should not be interpreted as a simple difference in acceptance of the system. Instead, men tended to frame limitations through realism and technical fidelity, while women more frequently described how these limitations affected control, comfort, and task execution.

5.4.3 VR-familiarity comment patterns

Participants with lower self-reported VR familiarity tended to describe the experiment as a discovery experience. Their comments often emphasized novelty, enjoyment, surprise, and the feeling of trying VR or simulated driving for the first time. In the VR condition, these participants generally focused less on technical details and more on immersion, the challenge of driving, and the potential use of the simulator for learning. Some low-familiarity participants also reported discomfort, especially nausea and dizziness, but these reports were usually embedded in broader comments about the experience being new, interesting, or fun. In the MR condition, lower-familiarity users often treated the real car as an important interpretive anchor. Seeing the physical vehicle appeared to support spatial understanding and made the experience easier to interpret, particularly for participants who were also inexperienced drivers.

Participants with higher VR familiarity produced more specific evaluations of calibration, interaction fidelity, and sensory consistency. In the VR condition, they were more likely to identify misalignment between the physical and virtual car, limited visual realism, pedal-sensitivity problems, absence of physical feedback, and mismatches between visual motion and the stationary body. Their comments suggest that prior exposure to immersive systems may have made them more attentive to implementation details that affected the credibility of the simulation. In the MR condition, higher-familiarity participants often evaluated how well the real vehicle and virtual elements were integrated. Some described the overlay as improving realism and control, while others noted that imperfect alignment or the reduction of digital cockpit elements made the experience less immersive. This indicates that VR familiarity did not simply increase satisfaction or criticism; rather, it changed the criteria used to evaluate the simulator, shifting attention from novelty toward calibration quality, perceptual coherence, and interaction fidelity.

5.4.4 Driving-experience comment patterns

Driving experience was strongly reflected in how participants interpreted the simulator. Licensed participants more often compared the XR vehicle with their expectations from real driving. In the VR condition, their comments frequently addressed the realism of acceleration, braking, steering, vehicle inertia, and the behavior of the car after releasing the accelerator. Several licensed participants described the braking response as too immediate or the accelerator as overly sensitive, and some highlighted that the lack of physical acceleration made it difficult to judge speed and movement. Licensed participants also noticed alignment problems between the real and virtual vehicle, especially when the body, mirrors, hands, or cockpit references did not coincide. These comments suggest that prior driving experience provided a reference model for evaluating the simulator, making mismatches in vehicle dynamics and spatial calibration more noticeable. In the MR condition, licensed participants often appreciated the real car as a reference for position and distance, but they still criticized curves, pedal behavior, speed cues, and residual desynchronization between the physical and virtual elements.

Unlicensed participants more frequently described the simulator through novelty, challenge, and learning potential. In the VR condition, several non-drivers emphasized that the experience was fun, immersive, or their first opportunity to interact with a driving-like system. Difficulties with the task were often attributed to not knowing how to drive, especially when participants reported leaving the track, struggling with curves, or needing time to understand the controls. In the MR condition, the visibility of the real car appeared particularly useful for some unlicensed participants because it improved orientation and helped them feel more control over the vehicle. Some comments explicitly suggested that the simulator could support introductory driving lessons or reduce fear of taking the wheel before real-world driving. However, unlicensed participants also reported pedal sensitivity, dizziness, nausea, and distraction caused by the collectible task. Taken together, the comments indicate that license status shaped not only performance, but also the meaning attributed to the simulator: licensed participants evaluated fidelity against real driving, while unlicensed participants tended to evaluate accessibility, learnability, and the experience of becoming familiar with vehicle control.

5.5 Cybersickness assessment

Cybersickness was examined as a subjective counterpart to the objective performance results, since the real-vehicle XR simulator required participants to coordinate physical vehicle controls with visually mediated motion cues. This assessment is relevant because discomfort in this context may be affected not only by the XR condition itself, but also by user factors such as gender and driver's license status. We summarize cybersickness using both individual symptom ratings and a composite score computed as the mean of the seven symptom ratings. Across the full sample, most scores remained close to the lower end of the 1–7 scale, suggesting that severe cybersickness was not generalized. Nevertheless, dizziness and nausea showed broader dispersion and higher values than the remaining symptoms, indicating that discomfort was concentrated mainly in vestibular and visual–vestibular responses.

Figure 10 shows cybersickness ratings by participant gender. Among men, cybersickness remained comparatively low across most symptoms, with visual fatigue, eye pain,

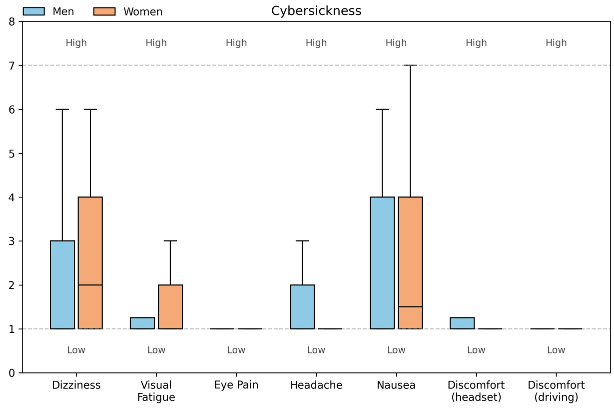

Figure 10. Cybersickness symptom ratings by participant gender. Symptoms were self-reported on a 1–7 scale, with higher values indicating greater discomfort.

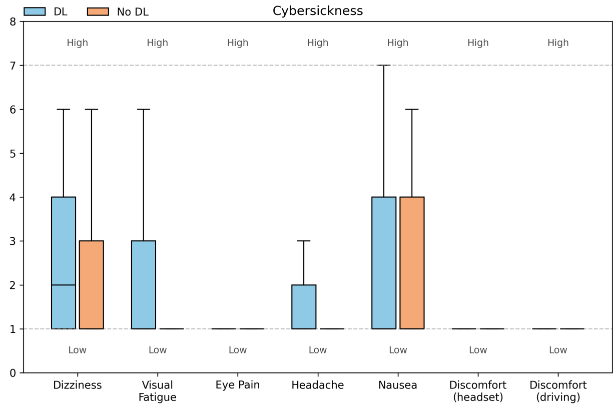

Figure 11. Cybersickness symptom ratings by driver's license status. Symptoms were self-reported on a 1–7 scale, with higher values indicating greater discomfort.

headache, headset discomfort, and driving discomfort close to the minimum score. Dizziness and nausea showed the clearest dispersion. In the round-progression analysis, the mean cybersickness score among men increased from 1.54 in the first round to 1.81 in the second round. This increase was mainly driven by nausea, which rose from 1.95 to 2.45; headache, from 1.35 to 1.85; dizziness, from 1.80 to 2.15; and visual fatigue, from 1.60 to 1.95. Eye pain and driving discomfort remained close to the minimum, while headset discomfort slightly decreased. These results suggest that, among men, repeated exposure produced some increase in discomfort, but this increase was limited in magnitude and concentrated in a small set of symptoms.

Among women, the cybersickness profile was more pronounced, especially for symptoms associated with dizziness, nausea, and visual fatigue. Women showed higher central values for dizziness and nausea than men, with broader distributions in both cases. In the round-progression analysis, women's mean cybersickness score increased from 1.52 in the first round to 2.03 in the second round, representing a stronger progression than that observed among men. The largest increase occurred in nausea, from 2.05 to 3.25, followed by visual fatigue, from 1.40 to 2.40, and dizziness, from 2.20 to 2.95. Headache also increased, although less sharply, while eye pain, headset discomfort, and driving discomfort remained comparatively low. This pattern suggests that women did not report uniformly higher discomfort across all items; rather, the difference was concentrated in symptoms more directly related to visual and vestibular strain.

Figure 11 compares cybersickness ratings between licensed and unlicensed participants. Licensed participants presented an important contrast between objective performance and subjective comfort. Although the performance analyses showed that driver's license status was associated with better task execution, the cybersickness results do not indicate lower discomfort among licensed participants. Instead, licensed participants showed slightly higher values for dizziness, visual fatigue, headache, nausea, and discomfort while driving. This tendency also appeared in the round-progression analysis: the mean cybersickness score among licensed participants increased from 1.59 in the first round to 2.09 in the second round. The most relevant increases occurred in visual fatigue, from 1.60 to 2.60; nausea, from 2.00 to 3.00; and dizziness, from 2.05 to 2.90. Headache also increased from 1.40 to 2.00.

This pattern suggests that licensed participants may have controlled the simulator more effectively while also being more sensitive to mismatches between expected real-car behavior and the mediated XR driving experience. Unlicensed participants showed a lower and more stable cybersickness profile. Their values remained close to the minimum for most symptoms, with nausea and dizziness again showing the clearest variation. Their mean cybersickness score increased from 1.48 in the first round to 1.74 in the second round, which was lower than the increase observed among licensed participants. The strongest change occurred in nausea, from 2.00 to 2.70, followed by visual fatigue, from 1.40 to 1.75, and dizziness, from 1.95 to 2.20. Headache, eye pain, headset discomfort, and driving discomfort remained near the low end of the scale. These results suggest that the absence of a driver's license was not associated with higher cybersickness in this sample, even though unlicensed participants showed more difficulty controlling the vehicle during the task.

Crossing gender and driver's license status provides a more nuanced interpretation of cybersickness in the real-vehicle XR simulator. The strongest increase from the first to the second round occurred among licensed women, whose mean cybersickness score rose from 1.39 to 2.10. This subgroup combined relatively low first-round discomfort with a marked second-round increase, especially in visual fatigue, dizziness, and nausea. Unlicensed women also showed an increase, but less sharply, from 1.66 to 1.96. Among men, licensed participants increased from 1.79 to 2.09, while unlicensed participants showed the smallest change, from 1.30 to 1.53.

Overall, the crossed analysis suggests that second-round discomfort was more concentrated among licensed participants, with licensed women showing the clearest progression. This finding separates cybersickness from objective driving performance: the profiles associated with better control and track adherence were not necessarily the profiles with lower discomfort. In this dataset, simulator performance and tolerance to XR-induced sensory strain should therefore be interpreted as complementary but distinct aspects of the user experience.

6 Discussion

The results indicate that objective performance in the real-vehicle XR simulator was shaped more strongly by participant background than by visualization condition alone. Although all participants completed the same track, task logic, physical vehicle controls, and VR/MR conditions, the main performance indicators did not respond uniformly to rendering mode. Batteries collected and inside-track percentage were largely stable across VR and MR, whereas first-lap duration was shorter in VR. This suggests that the fully virtual condition may have supported faster initial course progression, but not necessarily better driving quality or higher task success. In contrast, driver's license status, practical driving exposure, and recency of driving were more informative for explaining who controlled the simulator with greater stability. The central result of this study is therefore not that one XR condition was globally superior, but that embodied automotive XR evaluations must account for participants' prior relationship with vehicle control, since user background can strongly shape performance interpretation.

The gender-related results require cautious interpretation because the statistical pattern does not support a simple or intrinsic difference between women and men. In the mixed-effects models, the coefficient for men was significant for batteries collected ($\beta = 10.10, p < .001$) and inside-track percentage ($\beta = 15.26, p = .004$) relative to the reference subgroup. However, the subgroup analyses clarify that this pattern was concentrated in the unlicensed groups. Licensed women and licensed men did not differ significantly in either condition ($p = 1.000$), and, when only licensed drivers were considered, gender no longer predicted batteries collected ($p = .117$) or inside-track percentage ($p = .321$). The largest separation appeared where gender overlapped with lack of driving experience, while the licensed groups remained much closer in both battery collection and track adherence. The open-ended comments support this more nuanced interpretation: men tended to evaluate the simulator through technical fidelity, vehicle realism, pedal response, speed perception, and real-virtual alignment, whereas women more often emphasized control difficulty, discomfort, curve handling, and adaptation. These qualitative differences do not indicate that gender alone determined performance. Rather, they suggest that gender-related patterns in this sample were intertwined with practical exposure, confidence with driving-like tasks, discomfort sensitivity, and the criteria participants used to evaluate the simulator.

Driver's license status provided the clearest performance distinction across the study. Licensed participants collected more batteries and stayed inside the track for longer than unlicensed participants, with license status reaching significance for batteries collected ($\beta = 11.70, p < .001$) and inside-track percentage ($\beta = 19.03, p < .001$). The descriptive table and figures show that licensed men and licensed women achieved high and comparatively stable values in both VR and MR, while unlicensed participants, particularly unlicensed women, presented lower means and greater variability. This difference was also visible in the dispersion analysis: licensed participants showed lower variability in inside-track percentage, suggesting that prior driving exposure acted as a stabilizing factor when participants had to coordinate real pedals, steering, and visually mediated motion. The learning-rate analysis reinforces this interpretation. Unlicensed participants significantly reduced lap duration in the second session ($p = 0.034, d = 0.496$), suggesting adaptation to the controls, while licensed participants showed a ceiling effect, with nearly unchanged lap times ($p = 0.856, d = 0.032$) and less than 1% improvement in track accuracy.

The practical-experience analysis adds another layer to this result. Participants with no driving history differed significantly from those with any driving exposure in batteries collected ($H = 14.77, p < .001$) and inside-track percentage ($H = 15.09, p < .001$), while the two experienced groups were comparatively similar. Among licensed drivers, accumulated experience predicted slightly better battery collection and track adherence, but recency of driving was the stronger indicator: more months since last driving predicted fewer batteries ($\beta = -0.086, p < .001$) and less time inside the track ($\beta = -0.080, p < .001$). Thus, license status is useful because it captures a meaningful division in embodied vehicle-control competence, but it remains a coarse proxy. It becomes more informative when complemented by self-reported practical exposure and recent driving activity.

When track adherence and battery collection are examined jointly across VR and MR, the results indicate that gender differences should be interpreted through their overlap with driver's license status rather than as an isolated demographic effect. Condition-wise means showed that women had slightly higher battery collection and inside-track percentage in VR than in MR, whereas men showed slightly higher track adherence in MR than in VR. However, the four-level user-profile $\times$ condition interaction was not statistically significant ($p = 0.276$), and licensed-only analyses showed that gender no longer predicted batteries collected ($p = .117$) or inside-track percentage ($p = .321$). Therefore, the apparent tendency for women to perform better in VR and men to perform better in MR should be treated as descriptive. It may point to how embodiment, license status, and condition-specific affordances interact, but it does not support the claim that either gender has a stable advantage in one visualization mode.

Cybersickness followed a different pattern from objective task performance, showing that control effectiveness and sensory tolerance were related but not equivalent aspects of user experience. Across the sample, symptoms remained generally low, but dizziness, nausea, and visual fatigue were the most prominent sources of discomfort. Gender analysis showed that men had a smaller increase in mean cybersickness score from the first to the second round, rising from 1.54 to 1.81, while women increased from 1.52 to 2.03. This difference was not uniform across all symptoms; it was concentrated mainly in nausea, dizziness, and visual fatigue, which aligns with the qualitative comments in which women more often described physical discomfort and perceptual strain. License status produced an additional contrast between performance and comfort. Licensed participants controlled the simulator more effectively, but they did not report lower cybersickness. Their mean cybersickness score increased from 1.59 to 2.09, with larger changes in visual fatigue, nausea, and dizziness, whereas unlicensed participants increased from 1.48 to 1.74

and remained closer to the lower end of the scale. The crossed analysis further showed that licensed women presented the clearest progression, rising from 1.39 to 2.10, followed by licensed men, unlicensed women, and unlicensed men.

These results suggest that participants with stronger real-driving references may be more sensitive to mismatches between expected vehicle behavior, stationary body cues, and visually simulated motion, even when they perform well objectively. Consequently, cybersickness should not be treated as a secondary confirmation of performance quality. In real-vehicle XR simulators, performance metrics and discomfort metrics should be analyzed as complementary dimensions: the profiles that drive more accurately are not necessarily the profiles that tolerate XR-induced sensory conflict more comfortably.

7 Conclusion

This paper investigated the role of gender, driver's license status, and practical driving experience in objective performance within a real-vehicle XR simulator built around a stationary BYD vehicle. Across 80 logged runs from 40 participants, driver's license status provided the clearest and most consistent distinction in performance. Licensed participants generally collected more batteries, remained inside the track for longer, and showed lower variability than unlicensed participants. Visualization condition mainly affected first-lap duration, with shorter laps in VR, while batteries collected and inside-track percentage remained comparatively stable across VR and MR.

The gender-related results require a more cautious interpretation. Aggregate gender effects appeared in batteries collected and track adherence, but these effects were concentrated in the unlicensed subgroup and weakened substantially when the analysis focused on licensed drivers. Practical experience and driving recency further clarified this pattern: among licensed participants, recent driving activity was more informative than gender for explaining performance differences. These findings suggest that real-vehicle XR evaluations should profile participants carefully and avoid interpreting aggregate group differences too quickly as intrinsic properties of gender or display mode. In embodied automotive XR, practical exposure and recent driving activity may be more informative than demographic labels alone.

This study also has limitations. First, the sample was a convenience sample composed of participants linked to a computing-centered academic environment, so the results should not be generalized to the broader driving population without caution. Second, practical experience and driving recency were self-reported rather than independently verified. Third, the experience grouping was exploratory and partly shaped by the natural bins present in the collected responses. Fourth, the open-ended comments were used interpretively rather than through a formal qualitative coding procedure. Fifth, the task was intentionally abstracted into track following and battery collection, and therefore does not capture the full complexity of real traffic, social interaction, or road decision making.

Future work should expand the sample, include participants with more diverse driving backgrounds, add formal qualitative coding, investigate the relationship between head and hand positions and subjective responses, incorporate telemetry-derived control features, and explicitly test how calibration quality and feedback cues mediate the relationship between participant background and driving performance.

Declarations

Acknowledgements

The authors thank the volunteers who participated in the study and contributed their time to the evaluation of the real-vehicle XR simulator.

Funding

The authors have no funding to report for this research.

Authors' contributions

GCD: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft. **DOSC**: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft. **CGBP**: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft. **MHK**: Formal analysis, Methodology, Validation, Visualization, Writing – original draft. **WLC**: Conceptualization, Data curation, Formal analysis, Investigation, Project administration, Resources, Supervision, Writing – review & editing. **VT**: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **FLSNM**: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **JMT**: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Data and materials supporting this study are available upon request at `https://matched-realities-xr-driving.github.io/`.

References

Aldoski, Z. N. (2025). Virtual reality for autonomous vehicles: A review of safety, training, and human–machine interaction. *Dasinya Journal for Engineering and Informatics*, 1(1). DOI: 10.65542/djei.v1i1.8.

Bella, F. (2008). Driving simulator for speed research on two-lane rural roads. *Accident Analysis Prevention*, 40(3):1078–1087. DOI: 10.1016/j.aap.2007.10.015.

Colley, M., Jansen, P., Rukzio, E., and Gugenheimer, J. (2022). Swivr-car-seat: Exploring vehicle motion effects on interaction quality in virtual reality automated driving using a motorized swivel seat. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(4). DOI: 10.1145/3494968.

Darty, K., Saunier, J., and Sabouret, N. (2014). Effects of Gender on Presence and Virtual Driver Perception in Driving Simulators. In *5th International Conference on Women's Issues in Transportation*, page 19p, Paris - La Défense, France. Available at:`https://hal.science/hal-01025458`.

de Winkel, K. N., Talsma, T. M. W., and Happee, R. (2022). A meta-analysis of simulator sickness as a function of simulator fidelity. *Experimental Brain Research*, 240(12):3089–3105. DOI: 10.1007/s00221-022-06485-6.

Ghiurãu, F.-T., Baytaş, M. A., and Wickman, C. (2020). Arcar: On-road driving in mixed reality by volvo cars. In *Adjunct Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20 Adjunct, page 62–64, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3379350.3416186.

Goedicke, D., Bremers, A. W., Yasuda, H., and Ju, W. (2021). Xr-oom: Mixing virtual driving simulation with real cars and environments safely. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '21 Adjunct, page 67–70, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3473682.3480266.

Hassan, M., Shraban, S., Islam, M. A., Basaruddin, K., Ijaz, M., Kamarrudin, N., and Takemura, H. (2025). Integration of extended reality technologies in transportation systems: A bibliometric analysis and review of emerging trends, challenges, and future research. *Results in Engineering*, 26:105334. DOI: 10.1016/j.rineng.2025.105334.

Himmels, C., Rock, T., Venrooij, J., and Riener, A. (2022). Simulator fidelity influences the sense of presence in driving simulators. In *Adjunct Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '22, page 53–57, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3544999.3552526.

Hock, P., Benedikter, S., Gugenheimer, J., and Rukzio, E. (2017). Carvr: Enabling in-car virtual reality entertainment. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 4034–4044, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3025453.3025665.

Hock, P., Colley, M., Askari, A., Wagner, T., Baumann, M., and Rukzio, E. (2022). Introducing vampire – using kinaesthetic feedback in virtual reality for automated driving experiments. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '22, page 204–214, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3543174.3545252.

Hussain, Q., Alhajyaseen, W. K., Pirdavani, A., Reinolsmann, N., Brijs, K., and Brijs, T. (2019). Speed perception and actual speed in a driving simulator and real-world: A validation study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 62:637–650. DOI: 10.1016/j.trf.2019.02.019.

Kemeny, A. (2014). From driving simulation to virtual reality. In *Proceedings of the 2014 Virtual Reality International Conference*, VRIC '14, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/2617841.2620721.

Lønne, T. F., Karlsen, H. R., Langvik, E., and Saksvik-Lehouillier, I. (2023). The effect of immersion on sense of presence and affect when experiencing an educational scenario in virtual reality: A randomized controlled study. *Heliyon*, 9(6). DOI: 10.1016/j.heliyon.2023.e17196.

Mayhew, D. R., Simpson, H. M., Wood, K. M., Lonero, L., Clinton, K. M., and Johnson, A. G. (2011). On-road and simulated driving: Concurrent and discriminant validation. *Journal of Safety Research*, 42(4):267–275. DOI: 10.1016/j.jsr.2011.06.004.

McGill, M., Wilson, G., Medeiros, D., and Brewster, S. A. (2022). Passengxr: A low cost platform for any-car, multi-user, motion-based passenger xr experiences. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3526113.3545657.

Morra, L., Lamberti, F., Pratticó, F. G., Rosa, S. L., and Montuschi, P. (2019). Building trust in autonomous vehicles: Role of virtual reality driving simulators in hmi design. *IEEE Transactions on Vehicular Technology*, 68(10):9438–9450. DOI: 10.1109/TVT.2019.2933601.

Postelnicu, C.-C. and Boboc, R. G. (2024). Extended reality in the automotive sector: A bibliometric analysis of publications from 2012 to 2022. *Heliyon*, 10(2). DOI: 10.1016/j.heliyon.2024.e24960.

Schroeter, R. and Gerber, M. A. (2018). A low-cost vr-based automated driving simulator for rapid automotive ui prototyping. In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '18, page 248–251, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3239092.3267418.

Silvera, G., Biswas, A., and Admoni, H. (2022). Dreyevr: Democratizing virtual reality driving simulation for behavioural & interaction research. page 639–643. DOI: 10.5555/3523760.3523846.

Stahl, P., Donmez, B., and Jamieson, G. A. (2016). Supporting anticipation in driving through attentional and interpretational in-vehicle displays. *Accident Analysis Prevention*, 91:103–113. DOI: 10.1016/j.aap.2016.02.030.

Törnros, J. (1998). Driving behaviour in a real and a simulated road tunnel—a validation study. *Accident Analysis Prevention*, 30(4):497–503. DOI: 10.1016/S0001-4575(97)00099-7.

Winkler, I., Murari, T. B., Ferreira, C. V., and Freitas, F. (2022). Vr-based product development process: opportunities and challenges in the automotive industry. *Simpósio de*

Realidade Virtual e Aumentada (SVR), pages 41–44. DOI: 10.5753/svr$_e stendido$.2022.226711.