Journal of the Brazilian Computer Society, 2026, 32:1, doi: 10.5753/jbcs.2026.8243

Does immersion favor interaction? A within-subjects evaluation of desktop and HMD-based Virtual Reality interaction in an educational cartography application

Leticia Zanellatto de Oliveira [**Universidade Estadual do Oeste do Paraná-Unioeste** | ***leticia.oliveira21@unioeste.br***]
Matheus Prokopowiski dos Santos [**Universidade Estadual do Oeste do Paraná-Unioeste** | ***matheus.santos22@unioeste.br***]
Gustavo Camargo Domingues [**Universidade Federal de Pernambuco** | ***gcd2@cin.ufpe.br***]
Thaynara Nascimento de Jesus [**Universidade Estadual do Oeste do Paraná-Unioeste** | ***thaynara.jesus@unioeste.br***]
Claudio Roberto Marquetto Mauricio [**Universidade Estadual do Oeste do Paraná-Unioeste** | ***claudio.mauricio@unioeste.br***]
Ericson Hideki Hayakawa [**Universidade Estadual do Oeste do Paraná** | ***ericson.hayakawa@unioeste.br***]
João Marcelo Teixeira [**Universidade de São Paulo, Universidade Federal de Pernambuco**| ***jmxnt@cin.ufpe.br***]
Fatima L. S. Nunes [**Universidade de São Paulo** | ***fatima.nunes@usp.br***]
Fabiana Frata Furlan Peres [**Universidade de São Paulo, Universidade Estadual do Oeste do Paraná** | ***fabiana.peres@unioeste.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:** 05 August 2026 • **Published:** 14 August 2026

Abstract. Immersive Virtual Reality has been increasingly integrated into educational contexts due to its potential for spatial engagement. However, the trade-offs between immersion and interaction performance remain a subject of debate. This paper evaluates this relationship using GeoConnect, an educational cartography application designed to support Brazilian cartography learning through interactive map-assembly tasks. We conducted a within-subjects study with 36 students aged 12–13 years to compare two functionally equivalent 3D versions of GeoConnect, evaluated as a conventional Desktop condition and an HMD-based condition. The evaluation adopts a mixed subjective–objective approach, combining standardized instruments, SUS, UEQ, and SSQ, with telemetry-based metrics, including completion time, errors, precision, and a hesitation index. Results indicate that both conditions achieved high perceived usability, with mean SUS scores of 78.19 for the Desktop condition and 79.72 for the HMD-based condition. The HMD-based condition was preferred by 88.89% of participants, suggesting a strong preference for the immersive condition. However, cybersickness was not negligible: the HMD-based condition produced a mean SSQ total score of 32.41, a median of 22.44, and the highest subscale score in Disorientation; the most frequent symptoms were fullness of head, eyestrain, and dizziness with eyes open. Objective telemetry suggested that immersive interaction increased completion time by 82.97 seconds on average, although this difference did not survive Holm–Bonferroni correction across telemetry metrics. Presentation order also influenced completion time: the HMD time cost was concentrated among participants who used the HMD-based condition first, whereas participants who used Desktop first completed the later HMD-based condition with nearly equivalent times. These findings indicate descriptive pragmatic–hedonic tendencies moderated by learning and carryover effects. We conclude that Desktop and HMD-based conditions should be treated as complementary tools in geography education rather than mutually exclusive alternatives.

Keywords: Virtual Reality, Cartography Education, Human–Computer Interaction, User Experience.

1 Introduction

Virtual Reality (VR) has become an increasingly relevant medium for educational applications because it can combine interactive three-dimensional representations, stereoscopic visualization, head tracking, and embodied forms of manipulation. In contrast to conventional desktop interfaces, head-mounted displays (HMDs) can provide wide-field visual immersion and sensorimotor coupling, supporting the subjective sensation of being located inside a computer-generated environment [Slater, 2009; Jerald, 2015]. These characteristics are particularly relevant for learning tasks that involve spatial reasoning, object manipulation, and the understanding of relationships between parts and wholes, such as geography and cartography activities.

Despite this potential, the educational value of immersive VR cannot be assumed solely from its technological affordances. Meta-analytic evidence indicates that immersive system features, such as stereoscopy, tracking, and field of view, tend to increase presence, but their effects on task performance and learning outcomes are more dependent on the task, instructional design, and interaction demands [Cummings and Bailenson, 2016]. In education, previous reviews and meta-analyses suggest that VR-based instruction can support learning, motivation, and engagement, but they also emphasize that benefits vary across domains, interaction designs, and evaluation protocols [Merchant *et al.*, 2014; Radianti *et al.*, 2020]. More recent theoretical work further argues that immersive learning outcomes are mediated by factors such as presence, agency, interest, embodiment, cognitive load, and self-regulation [Makransky and Petersen, 2021]. Therefore, the central question is not simply whether immersive VR is more engaging than a desktop interface, but whether it provides measurable advantages without introducing excessive interaction cost, discomfort, or usability barriers.

This paper investigates these trade-offs between immersion and interaction through GeoConnect, a VR educational cartography application designed to support the learning of Brazilian cartography. In GeoConnect, users complete map-assembly tasks involving Brazilian macro-regions and states. The application was implemented in two functionally equivalent 3D versions: a Desktop version (non-immersive), operated through a conventional computer interface, and an HMD-based version, operated through immersive visualization and spatial interaction. Importantly, the comparison is not treated as a pure isolation of stereoscopy alone. Moving from desktop to HMD-based VR changes several interface characteristics simultaneously, including visual immersion, head tracking, field of view, input modality, and embodied manipulation. Thus, the study compares two realistic interaction configurations for the same educational task.

To evaluate these configurations, this work adopts a mixed subjective–objective approach. Subjective evaluation is based on standardized instruments widely used in human-computer interaction and VR research: the System Usability Scale (SUS) for perceived usability [Brooke *et al.*, 1996], the User Experience Questionnaire (UEQ) for pragmatic and hedonic dimensions of user experience [Laugwitz *et al.*, 2008], and the Simulator Sickness Questionnaire (SSQ) for symptoms associated with simulator and VR exposure [Kennedy *et al.*, 1993].

These instruments are complemented by telemetry logs collected during the task, including completion time, attempts, correct placements, errors, and precision. In addition, this work proposes a new telemetry-based metric, named the *hesitation index*, defined from the number of additional attempts beyond the expected correct placements. The proposed index was designed to capture interaction uncertainty, corrective behavior, repeated manipulations, and possible difficulties during spatial assembly tasks. This combination allows the analysis to distinguish between what users prefer, how they perceive each version, and how they actually perform during interaction. The study is guided by the following research questions:

- **RQ1:** How do the desktop and HMD-based conditions of GeoConnect differ in perceived usability?
- **RQ2:** How do the two conditions differ across pragmatic and hedonic dimensions of user experience?
- **RQ3:** How do the two conditions differ in objective task performance, considering completion time, attempts, errors, precision, and hesitation?
- **RQ4:** What level of cybersickness is reported after the HMD-based condition?
- **RQ5:** How do user preference and qualitative feedback help explain the subjective and objective results?
- **RQ6:** Did the order in which participants experienced the Desktop and HMD-based conditions influence subjective or objective outcomes?

The contributions of this paper are fourfold. First, it presents a controlled comparison between desktop and HMD-based interaction in the same educational cartography application, reducing confounds related to different learning contents or task structures. Second, it combines standardized subjective instruments with telemetry-based objective metrics, enabling a more complete interpretation of usability, experience, and performance. Third, the paper proposes a new telemetry-derived metric, the *hesitation index*, intended to quantify interaction fluency and manipulation uncertainty during spatial assembly tasks. Fourth, it discusses design implications for educational VR applications that require spatial manipulation, particularly when immersive interaction must be balanced against usability, comfort, and classroom feasibility.

The remainder of this paper is organized as follows. Section 2 reviews related work on immersive VR in education, desktop-versus-HMD interaction, spatial learning, geography education, and telemetry-based evaluation. Section 3 presents GeoConnect, the evaluated conditions, participants, procedure, instruments, telemetry processing, and statistical analysis. Section 4 reports the subjective and objective results. Section 5 discusses the findings, design implications, and threats to validity. Finally, Section 6 concludes the paper and outlines future work.

2 Related Work

This section reviews previous work on immersive virtual reality in education, comparisons between desktop and HMD-based interfaces, spatial learning in virtual environments, and subjective–objective evaluation methods for VR educational cartography applications. These four perspectives are directly related to the present study because GeoConnect combines an educational spatial task, two interface conditions, standardized questionnaires, and telemetry-based performance measures.

2.1 Immersive Virtual Reality in Education

Immersive virtual reality has been widely investigated as a medium for educational applications because it can combine three-dimensional visualization, interaction, embodiment, and a strong sense of presence. Dede [2009] argues that immersive interfaces can support learning through experiential engagement, situated interaction, and forms of representation that are difficult to reproduce with conventional media. Similarly, meta-analytic and review studies suggest that VR-based instruction can improve engagement and support learning in several domains, although its effectiveness depends on instructional design, interaction quality, domain characteristics, and assessment method [Merchant *et al.*, 2014; Radianti *et al.*, 2020].

A central issue in this literature is that immersion is not automatically equivalent to better learning or better task performance. The Cognitive Affective Model of Immersive Learning (CAMIL) proposed by Makransky and Petersen [2021] argues that immersive VR affects learning through psychological affordances such as presence and agency, which in turn influence factors such as interest, intrinsic motivation, self-efficacy, embodiment, cognitive load, and self-regulation.

This perspective is important for educational VR evaluation because it suggests that a highly immersive interface may increase motivation and perceived engagement while also increasing cognitive or interaction demands. Therefore, studies comparing immersive and non-immersive interfaces should examine not only user preference, but also usability, task efficiency, discomfort, and objective performance.

Recent work has also moved CAMIL-informed immersive learning discussions closer to realistic educational deployments. For example, Kablitz [2025] used CAMIL-related constructs in a quasi-experimental field study with vocational-school trainees, showing that immersive VR can support knowledge acquisition and transfer-related self-efficacy while also requiring attention to motivational, self-regulatory, and transfer processes. Similarly, Liu *et al*. [2025] investigated immersive VR in a classroom-oriented educational setting and found that interactive VR produced stronger student experience, concentration, and knowledge-mastery outcomes than non-interactive VR or traditional video instruction. These studies reinforce the view that immersive learning should not be evaluated only through technological immersion, but through the interaction between presence, agency, motivation, cognitive demands, usability, and instructional context. Therefore, in the present study, CAMIL is used not as a claim that immersion should automatically improve learning, but as a framework for interpreting why an HMD-based interface may increase preference and engagement while also introducing interaction demands.

2.2 Desktop versus HMD-Based Learning Environments

Several studies have compared immersive VR with less immersive desktop-based alternatives. Parong and Mayer [2018] compared a biology lesson presented in immersive VR with an equivalent desktop slideshow. Their results showed that immersive VR increased presence and affective engagement, but learning benefits were not automatic and depended on the use of appropriate learning strategies. This result is particularly relevant to the present study because it illustrates a common trade-off in educational VR: immersion may improve subjective experience, but objective outcomes may depend on instructional and interaction factors.

Makransky and Lilleholt [2018] also investigated the emotional value of immersive VR in education, showing that immersive VR can increase presence and affective value, but that these effects must be interpreted together with learning and usability outcomes. In complementary work, Makransky and Petersen [2019] analyzed desktop VR learning and identified both affective and cognitive paths through which VR features, presence, usability, and cognitive benefits can influence learning outcomes. Together, these studies suggest that desktop and HMD-based interfaces should not be understood simply as weaker or stronger versions of the same medium. Instead, each condition offers different combinations of usability, familiarity, interaction demands, and affective engagement.

Classic work on spatial knowledge acquisition in virtual environments remains important because it established that display configuration, viewpoint control, and locomotion can influence navigation and spatial understanding. However, those findings should be interpreted as foundational rather than directly representative of current standalone HMDs. Modern HMDs provide higher-resolution displays, lower latency, inside-out tracking, wider consumer availability, and more naturalistic head/body coupling than earlier systems. Recent spatial-cognition work therefore emphasizes that spatial learning in immersive VR depends not only on stereoscopy, but also on locomotion mode, embodied movement, field of view, tracking fidelity, interaction technique, and task design. This is directly relevant to GeoConnect because participants do not merely observe a map: they manipulate pieces, inspect spatial relations, and correct placements through feedback.

2.3 Spatial Learning, Cartography, and Educational Applications

Spatial learning is a recurring motivation for using interactive three-dimensional media in education. Geography and cartography tasks often require learners to understand part–whole relations, relative position, scale, orientation, and symbolic representation. Previous work with augmented and virtual environments has explored how interactive spatial representations can support these forms of understanding. For example, Shelton and Hedley [2002] used augmented reality to teach Earth–Sun relationships to geography students, emphasizing the role of interactive manipulation in understanding spatial phenomena. Carbonell-Carrera and Saorin [2017] investigated virtual learning environments for spatial orientation, reinforcing the relevance of virtual environments for geospatial and orientation-related skills.

More recent geography-education research has extended immersive VR beyond early AR demonstrations toward school-based geography lessons, virtual field trips, collaborative immersive environments, and geospatial learning activities. Czimre *et al*. [2024], in a systematic review of VR for geography teaching, show that the field has increasingly examined not only motivation and engagement, but also classroom feasibility, hardware and software constraints, research methods, and the slow integration of VR into curricular geography education. Edler *et al*. [2025] describe an immersive VR learning environment for school geography in North Rhine-Westphalia, focused on smart-city topics and designed to support spatial reasoning, critical reflection, and decision-making through problem-oriented learning tasks. Střechová *et al*. [2025] further demonstrate the relevance of collaborative immersive virtual environments for geography education through a climate-zones case study with students aged 12–15 using Meta Quest 2 headsets, reporting enjoyment and perceived learning benefits alongside manipulation and interaction challenges. GeoConnect differs from these initiatives by focusing on a controlled within-subjects comparison between two functionally equivalent versions of the same Brazilian cartography task, combining standardized subjective instruments with telemetry-based interaction measures.

Educational games can support learning by combining goals, feedback, challenge, and interaction, but their effectiveness depends on how game mechanics align with learning objectives [Mayer, 2019]. In the context of GeoConnect, the educational objective is not simply to expose students to geographic information, but to engage them in an active map-assembly task. This makes the interface comparison particularly important: desktop interaction may offer conventional and potentially efficient manipulation, whereas HMD-based interaction may provide stronger spatial involvement, depth perception, and embodied engagement. The comparison therefore concerns both educational experience and interaction performance.

This perspective is aligned with cartographic literacy approaches that emphasize active manipulation, spatial orientation, and the understanding of part-whole geographic relations as fundamental processes in the development of spatial reasoning and map comprehension [Oliveira, 2011; Simielli, 2011].

2.4 Telemetry, Event Logs, and Interaction Fluency in VR

Questionnaire-based measures are important for evaluating perceived usability and user experience, but they do not fully describe how users interact with a system during task execution. For this reason, HCI and VR studies increasingly complement self-report instruments with objective measures, event logs, interaction traces, and automated indicators of usability problems. Harms [2019], for example, proposed an automated usability-evaluation approach for VR applications based on recordings of actual use, from which task trees are generated and analyzed to detect usability smells. Kamińska *et al.* [2022] similarly explored the automation of VR usability testing through objective features collected during interaction. More recently, Chen *et al.* [2024] reviewed VR interaction-interface evaluation methods and showed that objective evaluation commonly includes task-completion time, accuracy, errors, pauses, requests for assistance, revisiting steps, unnecessary actions, and other observable behaviors. In the present study, the hesitation index follows this logic. It does not measure learning directly. Instead, it operationalizes interaction fluency during the map-assembly task by counting additional placement attempts beyond the expected correct placements. Repeated attempts may indicate uncertainty, corrective behavior, manipulation difficulty, or difficulty identifying the correct target. Therefore, the hesitation index is interpreted as a telemetry-derived indicator of interaction uncertainty rather than as an educational achievement score.

2.5 Evaluating User Experience, Usability, and Cybersickness in VR

Evaluation of educational VR systems requires instruments that capture more than one dimension of the experience. The System Usability Scale is widely used as a compact measure of perceived usability [Brooke *et al.*, 1996]. However, usability alone does not capture broader user-experience qualities such as stimulation, novelty, attractiveness, and perceived control. The User Experience Questionnaire was designed to measure both pragmatic and hedonic dimensions of user experience [Laugwitz *et al.*, 2008], and its benchmark has been used to support interpretation of UX scores across interactive products [Schrepp *et al.*, 2017].

For HMD-based VR, discomfort and cybersickness are also central evaluation dimensions. The Simulator Sickness Questionnaire remains one of the most widely used instruments for quantifying nausea, oculomotor symptoms, disorientation, and total sickness score [Kennedy *et al.*, 1993]. Nevertheless, previous work distinguishes cybersickness from simulator sickness and highlights that symptom profiles may differ across systems and users [Stanney *et al.*, 1997]. More recent methodological discussions also warn that SSQ results in VR studies can be difficult to compare across protocols and should be reported carefully, preferably with subscales and sufficient contextual information [Bimberg *et al.*, 2020]. Reviews of immersive VR ergonomics further show that cybersickness, visual fatigue, muscular fatigue, stress, and mental load may affect the feasibility of replacing or complementing desktop interfaces with immersive systems [Souchet *et al.*, 2023].

Because of these concerns, subjective measures should be complemented with objective interaction data whenever possible. Telemetry-based metrics such as completion time, number of attempts, errors, and precision can reveal whether users who report a positive experience also perform efficiently and accurately. This is particularly important in educational applications, where preference and novelty effects may not align with task performance. The present study therefore combines SUS, UEQ, SSQ, post-experience preference, and telemetry logs to compare desktop and HMD-based versions of GeoConnect.

2.6 Positioning of the Present Study

The studies reviewed above show that immersive VR can increase presence, motivation, and engagement, but that its effects on objective performance and learning-related tasks are not uniform. Prior comparisons between desktop and HMD-based systems often focus either on learning outcomes, spatial navigation, or subjective experience. Fewer studies combine standardized usability and UX instruments with telemetry-based objective performance in a structurally equivalent educational application. The present work addresses this gap by evaluating (with different instruments) two versions of the same cartography game, preserving the same educational task while varying the interaction and display configuration. This design allows us to examine whether the HMD-based version provides advantages in user experience, preference, and spatial task performance, and whether these potential benefits are accompanied by usability costs or cybersickness symptoms.

3 Methods

This section describes GeoConnect, the evaluated interaction conditions, the participant sample, the within-subjects protocol, the subjective instruments, and the telemetry-based analysis. The purpose of this organization is to present the application and the experiment as a single methodological unit, avoiding an overly descriptive separation between system description, study design, and data processing.

3.1 GeoConnect Application and Experimental Conditions

GeoConnect is an educational cartography application designed to support Brazilian cartography learning through interactive 3D map-assembly tasks. Participants manipulate map pieces and place them in their corresponding positions, progressively constructing representations of Brazilian macro-regions and states. The activity requires visual search, spatial reasoning, object manipulation, and feedback-based correction.

GeoConnect was implemented in two functionally equivalent software versions: a *Desktop version* and an *HMD-based version*. In the experiment, these two implementations were evaluated as two interaction conditions: the *Desktop condition* and the *HMD-based condition*. Both conditions preserved the same educational content, task sequence, correct answers, validation logic, tutorial structure, and telemetry variables, while differing in display technology, input device, visual immersion, and manipulation style.

The task was organized into six map boards. Participants first assembled the Brazilian macro-regions and then completed the state-placement boards for the North, Northeast, Midwest, Southeast, and South regions. In both conditions, each board contained a set of target sockets and a set of corresponding map pieces. A placement was accepted only when the released piece matched the expected target according to the validation rules. Correct placements were visually consolidated through the socket mechanism, whereas incorrect placements were rejected and required additional manipulation. Thus, the core puzzle logic was shared across conditions.

In the HMD-based condition, participants used a Meta Quest 3 headset and tracked controllers. The activity was performed standing inside a delimited safe area monitored by the experimenter. Continuous locomotion was controlled through the left-controller joystick using the Unity XR Interaction Toolkit movement action, while object selection and manipulation were performed with controller trigger/grip actions. Participants grabbed, moved, and released map pieces in 3D space. Because continuous artificial locomotion, body posture, and controller-based manipulation can affect ergonomics and cybersickness, these characteristics are considered when interpreting SSQ results.

In the Desktop condition, participants interacted with the same 3D environment using a conventional monitor, keyboard, and mouse. Movement was controlled with the keyboard, camera orientation with the mouse, and object selection through a ray-casting interaction from the camera viewpoint. When a piece was selected, the interaction alternated between a held and released state; placement was accepted only when the released piece satisfied the corresponding proximity and validation rules. Therefore, Desktop interaction was not treated as simple 2D dragging on a screen plane. Instead, participants navigated and manipulated objects in a 3D environment through a non-stereoscopic interface, relying on perspective, viewpoint control, occlusion, shading, and validation feedback for spatial interpretation.

This equivalence does not mean that the two conditions isolate stereoscopy alone. Moving from Desktop to HMD-based interaction also changes field of view, body involvement, locomotion, input modality, and spatial manipulation style. Therefore, the comparison should be interpreted as a realistic contrast between two complete interaction configurations for the same educational cartography task, rather than as a pure manipulation of stereoscopy.

3.2 Participants

The study included 36 students aged 12–13 years. The sample included 19 male participants and 17 female participants. Regarding prior HMD exposure, 16 participants reported never having used an HMD, 10 reported having used an HMD once, and 10 reported having used an HMD more than once but rarely. Thus, 20 participants had at least some previous contact with HMD-based virtual reality, whereas 16 had no previous HMD experience.

The characterization questionnaire recorded prior HMD experience, but it did not separately quantify everyday use of desktop/laptop computers, mouse/keyboard interaction, or mobile devices. Therefore, this paper avoids treating the Desktop condition as inherently familiar for all participants. Any interpretation involving desktop familiarity should be read cautiously, especially because students in this age range may be more accustomed to mobile devices than to conventional mouse-and-keyboard interaction.

Participants were recruited from a school context and took part in the study only after the required ethical documents had been completed. Because the participants were minors, legal guardians signed an informed consent form, and the students signed an assent form indicating their willingness to participate. An image and voice authorization form was also collected before data collection.

3.3 Study Design and Procedure

The study used a within-subjects design with two interaction conditions: the *Desktop condition* and the *HMD-based condition*. Each participant completed the same GeoConnect task in both conditions. The within-subjects design was chosen to reduce between-participant variability and to allow direct comparison of subjective and objective measures for the same participants.

To mitigate order effects, the presentation order was alternated across participants. Half of the participants completed the HMD-based condition first and then the Desktop condition, whereas the other half completed the Desktop condition first and then the HMD-based condition. This counterbalancing strategy allowed the analysis to examine whether task-learning, familiarization, or carryover effects influenced the observed differences between conditions.

The data collection took place in a reserved room configured to reduce external distractions and ensure physical safety during HMD use. Although the room layout allowed more than one participant to complete the experiment during the same collection period, participants were placed at separated stations, desktop screens were oriented to prevent one participant from observing another participant's task, and questionnaires were completed individually after each condition. Participants were instructed not to communicate

Table 1. Functional equivalence and interaction differences between the two GeoConnect versions.

Feature	Desktop version	HMD-based version
Display	Conventional monitor	Meta Quest 3 head-mounted display
Visual presentation	Non-stereoscopic 3D scene	Stereoscopic immersive 3D scene
Primary input	Mouse and keyboard	6-DoF tracked controllers
Navigation	Keyboard movement and mouse-controlled viewpoint	Continuous joystick-based locomotion
Selection	Camera-based ray casting and mouse click	Controller-based grabbing
Manipulation	Held/released interaction mediated by mouse and camera ray	Grab, move, and release in 3D space
Depth cues	Viewpoint navigation, perspective, occlusion, shading, and socket validation	Stereoscopy, head tracking, embodied manipulation, and socket validation
Tutorial	Shared structure adapted to mouse/keyboard controls	Shared structure adapted to HMD/controller controls
Task sequence	Macro-regions, North, Northeast, Midwest, Southeast, South	Macro-regions, North, Northeast, Midwest, Southeast, South
Puzzle logic	Shared	Shared
Validation rules	Shared	Shared
Telemetry variables	Shared	Shared

Table 2. Participant profile used in the paired analysis.

Variable	Result
Participants	36
Age	12–13 years
Male	19 participants
Female	17 participants
Never used HMD	16 participants
Used HMD once	10 participants
Used HMD more than once, rarely	10 participants
HMD-first order	18 participants
Desktop-first order	18 participants

during task execution, and an experimenter monitored the sessions. These procedures were intended to reduce visual and social interference. Nevertheless, because participants were not tested in fully isolated booths, residual peer-presence or social-contagion effects cannot be completely ruled out and are acknowledged as a threat to validity.

For each condition, the protocol followed the same general sequence: familiarization, task execution, and questionnaire completion. Figure 1 summarizes the experimental flow.

During the familiarization phase, participants explored the interface for approximately five minutes. This stage served as a tutorial and allowed participants to practice the basic interaction mechanics before the main task. In the Desktop condition, participants practiced keyboard movement, mouse-controlled viewpoint changes, selection, and release of pieces. In the HMD-based condition, participants practiced locomotion, grabbing, releasing, and positioning objects using the VR controllers.

After familiarization, participants performed the main map-assembly task. The same educational content, stage sequence, and validation logic were used in both conditions. Participants first assembled the Brazilian macro-regions and then completed the state-placement boards for the North, Northeast, Midwest, Southeast, and South regions. The application automatically recorded stage-level telemetry during task execution.

Between conditions, participants were given a short interval of approximately three to five minutes. This interval was included to reduce fatigue and potential carryover effects, especially after HMD use. After completing each condition, participants answered the subjective questionnaires associated with that condition. At the end of the complete experience, participants answered a post-experience questionnaire reporting their preferred condition and optional comments about the activity.

3.4 Measures and Instruments

The evaluation combined standardized subjective instruments, post-experience preference, open-ended feedback, and objective telemetry logs. SUS and UEQ were applied after both the Desktop and HMD-based conditions. SSQ was applied only after the HMD-based condition, because the purpose of the instrument in this study was to characterize symptoms associated with immersive headset use.

The System Usability Scale (SUS) was used to measure perceived usability. SUS consists of ten Likert-scale items and produces a score from 0 to 100, where higher values indicate better perceived usability [Brooke *et al*., 1996]. In this study, SUS was used to compare whether the HMD-based condition preserved, improved, or reduced perceived usability relative to the Desktop condition.

The User Experience Questionnaire (UEQ) was used to evaluate broader user-experience qualities. UEQ consists of 26 semantic differential items grouped into six scales: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty [Laugwitz *et al*., 2008]. These scales allow the comparison between pragmatic qualities, such as efficiency and clarity, and hedonic qualities, such as stimulation and novelty. This distinction is important because an HMD-based condition may be more engaging while not necessarily being more efficient.

The Simulator Sickness Questionnaire (SSQ) was used to assess cybersickness-related symptoms after HMD use. SSQ supports the calculation of Nausea, Oculomotor, Disorientation, and Total scores [Kennedy *et al*., 1993]. In the present study, SSQ was not used as a comparison between Desktop and HMD-based conditions, but as a comfort and safety measure for the immersive condition.

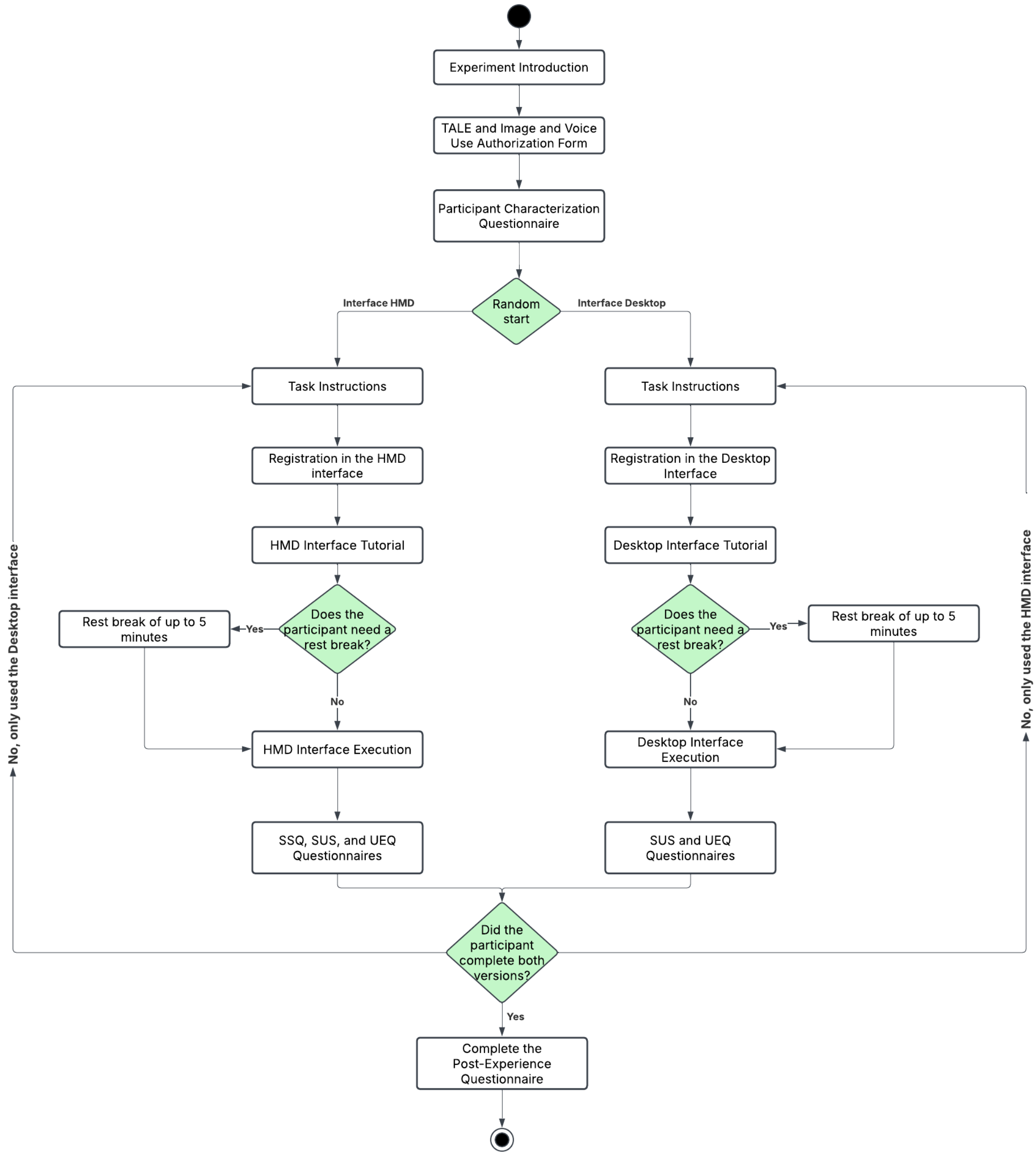

Figure 1. Experimental procedure followed during data collection.

Finally, the post-experience questionnaire collected participants' preferred condition and open-ended comments. These qualitative responses were used to contextualize the quantitative results and to identify design issues perceived by participants, such as clarity of feedback, ease of manipulation, visual legibility, and suggestions for improving the learning experience.

3.5 Telemetry Data Processing and Statistical Analysis

The dataset combined three sources: the questionnaire workbook, the HMD telemetry log, and the Desktop telemetry log. The questionnaire workbook contained participant characterization, SUS responses, UEQ responses, SSQ responses, post-experience preference, and open-ended feedback. The telemetry logs contained session-level identifiers and stage-level records for each condition of GeoConnect.

Participant identifiers were used only to match questionnaire responses and telemetry sessions across conditions. Since the telemetry files used independent session identifiers for each platform, participant matching was performed using the participant name or identifier, date of collection, and session completeness. Before matching, names were normalized by converting text to lowercase, removing accents, trimming extra spaces, and removing non-informative special characters introduced during data entry. After matching, participants were assigned anonymized study codes for analysis and reporting.

A participant was included in the main paired analysis only if all of the following criteria were satisfied:

1. the participant appeared in the questionnaire workbook;
2. the participant had one valid HMD telemetry session;
3. the participant had one valid Desktop telemetry session;
4. both telemetry sessions contained the expected sequence of GeoConnect stages;
5. both telemetry sessions contained valid duration, attempt, hit, and error values;
6. the session was not a debug, training-only, duplicated, or incomplete record.

Sessions with fewer than the expected stages, zero or missing stage durations, debug identifiers, or ambiguous duplicate records were excluded from the paired analysis. When multiple complete sessions were available for the same participant and condition, the session temporally consistent with the experimental protocol was retained. Ambiguous cases that could not be resolved reliably were excluded to avoid contaminating paired comparisons. After cleaning and matching, the final dataset contained 36 participants with valid questionnaire responses and complete paired telemetry sessions.

Telemetry was processed at the stage level rather than relying on final summary fields. This decision was necessary because some final-session fields were missing, zero-valued, or inconsistent across platform logs, whereas the stage-level records consistently stored the variables required for analysis. Each complete session contained the following six stages:

1. *Map Board 1 Start*;
2. *Map Board 2 North Start*;
3. *Map Board 2 North East Start*;
4. *Map Board 2 Midwest Start*;
5. *Map Board 2 South East Start*;
6. *Map Board 2 South Start*.

For each participant p, condition $c \in \{\text{Desktop}, \text{HMD}\}$, and stage s, the telemetry log provided duration in seconds (d_{pcs}), number of attempts (n_{pcs}), number of correct placements (h_{pcs}), and number of errors (e_{pcs}). Global session metrics were reconstructed by summing the corresponding stage-level values:

$$T_{pc} = \sum_{s=1}^{S} d_{pcs}, \tag{1}$$

$$N_{pc} = \sum_{s=1}^{S} n_{pcs}, \tag{2}$$

$$H_{pc} = \sum_{s=1}^{S} h_{pcs}, \tag{3}$$

$$E_{pc} = \sum_{s=1}^{S} e_{pcs}, \tag{4}$$

where T_{pc} is total completion time, N_{pc} is the total number of attempts, H_{pc} is the total number of correct placements, E_{pc} is the total number of errors, and S is the number of completed stages.

Two derived performance metrics were computed. First, precision was defined as the proportion of correct placements among all attempts:

$$P_{pc} = \frac{H_{pc}}{N_{pc}}. \tag{5}$$

Second, the hesitation index was defined as the number of additional attempts beyond the number of correct placements:

$$I_{pc}^{hes} = N_{pc} - H_{pc}. \tag{6}$$

The hesitation index captures repeated actions, uncertainty, failed manipulations, or corrective attempts. Because the task structure was equivalent across conditions and the expected number of correct placements was fixed, higher values indicate lower interaction fluency. The same aggregation procedure was also applied at the stage level, enabling exploratory comparisons of whether performance differences between Desktop and HMD-based interaction were concentrated in specific map boards.

SUS, UEQ, and SSQ responses were scored according to their standard procedures. For SUS, each item was scored from 1 to 5. Odd-numbered items were transformed by subtracting 1 from the participant response, while even-numbered items were transformed by subtracting the response from 5. The transformed values were summed and multiplied by 2.5, producing a score from 0 to 100:

$$SUS = 2.5 \left(\sum_{i \in O} (x_i - 1) + \sum_{i \in E} (5 - x_i) \right), \tag{7}$$

Table 3. Subjective and qualitative measures used in the study.

Construct	Instrument	Applied to	Outcome
Usability	SUS	Desktop and HMD-based conditions	0–100 usability score
User experience	UEQ	Desktop and HMD-based conditions	Six UX scale scores
Cybersickness	SSQ	HMD-based condition only	Nausea, Oculomotor, Disorientation, and Total scores
Preference	Post-experience questionnaire	After both conditions	Preferred condition
Qualitative feedback	Open-ended question	After both conditions	Comments and suggestions

where O is the set of odd-numbered items, E is the set of even-numbered items, and x_i is the response to item i.

UEQ responses were transformed to the standard scale ranging from -3 to $+3$, respecting the polarity of each semantic differential item. The six UEQ scales were then computed as the mean of their corresponding items: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. Scale-level scores were used for analysis rather than individual items.

SSQ was scored only for the HMD-based condition. Symptom responses were converted to numeric values and grouped into the standard Nausea, Oculomotor, and Disorientation subscales. Subscale and total scores were computed using the standard SSQ weighting procedure:

$$SSQ_{Nausea} = 9.54 \times N_{raw}, \tag{8}$$

$$SSQ_{Oculomotor} = 7.58 \times O_{raw}, \tag{9}$$

$$SSQ_{Disorientation} = 13.92 \times D_{raw}, \tag{10}$$

$$SSQ_{Total} = 3.74 \times (N_{raw} + O_{raw} + D_{raw}), \tag{11}$$

where N_{raw}, O_{raw}, and D_{raw} are the raw sums for each SSQ symptom group.

Missing questionnaire items were not imputed. If a participant had incomplete responses for a scale, that scale was excluded from the corresponding analysis. The participant could still be retained for other analyses if the required data were complete.

The main analysis compared Desktop and HMD-based conditions using paired participant-level values. For each metric X, a paired difference was computed as:

$$\Delta X_p = X_{p,HMD} - X_{p,Desktop}. \tag{12}$$

Positive values therefore indicate higher values in the HMD-based condition, while negative values indicate higher values in the Desktop condition. Completion time, attempts, errors, precision, hesitation index, SUS, and UEQ scales were analyzed using paired comparisons. SSQ was analyzed descriptively because it was applied only after the HMD-based condition.

Because the sample size was moderate and several variables were non-normally distributed, the Wilcoxon signed-rank test was used for paired comparisons between Desktop and HMD-based conditions. Effect size was reported using the rank-biserial correlation, r_{rb}, when applicable. For preference, the distribution of preferred condition was compared against an equal-preference baseline using an exact binomial test.

To control the family-wise error rate, Holm–Bonferroni correction was applied within conceptually related families of tests. SUS was treated as a single usability outcome. UEQ scales were corrected within the UEQ family. Global telemetry metrics were corrected within the telemetry family. Stage-level analyses were treated as exploratory and corrected within the stage-level family when formal comparisons were reported. Corrected and uncorrected results were distinguished in the Results section.

Telemetry distributions were inspected for extreme values, especially completion time, attempts, errors, and hesitation index. Extreme values were not removed solely because they were large, since unusually long completion times or many attempts may reflect genuine interaction difficulty. A record was excluded only when there was evidence of technical failure, incomplete gameplay, duplicated logging, debug use, or mismatch with questionnaire data. To assess robustness, sensitivity analyses were conducted by repeating the main paired comparisons after excluding extreme telemetry observations identified by the interquartile-range criterion. A result was considered robust when the direction and interpretation of the effect remained consistent across the full and sensitivity analyses. From the initial 42 participants, data from 6 of them were removed due to unmatched questionnaire or telemetry records (no valid paired record across sources), resulting in a total of 36 retained paired participants (with complete questionnaire, desktop telemetry and HMD telemetry).

Order effects were analyzed because each participant completed both conditions and could therefore benefit from task familiarization in the second condition. Participants were grouped according to their first condition: HMD-first or Desktop-first. For each paired metric, the difference score $\Delta X_p = X_{p,HMD} - X_{p,Desktop}$ was compared between order groups using the Mann–Whitney U test. In addition, first-session and second-session completion times were compared using the Wilcoxon signed-rank test to assess learning or carryover effects independently of interface condition. Preference distributions by order group were compared using Fisher's exact test. These order analyses were treated as exploratory because the study was primarily designed as a paired comparison between interface conditions.

All analyses used a significance level of $\alpha = 0.05$. However, interpretation emphasized effect sizes, confidence intervals, corrected p values, and consistency between subjective and objective measures rather than relying exclusively on uncorrected statistical significance. This choice is appropriate because the study focuses on interaction trade-offs between two educational interfaces, where practical differences in usability, comfort, preference, and performance may be relevant even when statistical power is limited.

4 Results

This section reports the results of the subjective questionnaires, post-experience preference, and telemetry-based performance analysis. Questionnaire results were computed from the new questionnaire workbook, while objective metrics were reconstructed from the stage-level telemetry logs. As described in Section 3.5, telemetry values were aggregated from the six expected GeoConnect stages because the stage-level records consistently store duration, attempts, hits, and errors for both conditions. The HMD telemetry log and the Desktop telemetry log both contain per-stage fields for *attempts*, *completionTime*, *correctPlacements*, and *errors*. The experimental protocol applies SUS and UEQ after both conditions, while SSQ was applied after the HMD-based condition, and a post-experience questionnaire collected preference and comments Brooke *et al*. [1996]; Laugwitz *et al*. [2008]; Kennedy *et al*. [1993].

4.1 Participant Profile

The questionnaire dataset contained valid responses from 36 participants. All participants were students between 12 and 13 years old. The sample included 19 male participants and 17 female participants. Prior HMD exposure was mixed: 20 participants reported having used an HMD at least once, whereas 16 reported no previous HMD use. Table 4 summarizes the participant profile.

Table 4. Participant profile based on the questionnaire dataset.

Variable	Result
Participants	36
Age	12–13 years
Male	19 participants
Female	17 participants
Never used HMD	16 participants
Used HMD once	10 participants
Used HMD more than once, rarely	10 participants
HMD-first order	18 participants
Desktop-first order	18 participants

The dataset contained valid questionnaire responses and complete paired telemetry sessions for 36 participants. Therefore, both subjective results and objective paired comparisons are reported with $n = 36$.

4.2 Perceived Usability

Table 5 presents the SUS scores for the Desktop and HMD-based conditions. Both conditions obtained mean scores above 75, indicating positive perceived usability. The HMD-based condition obtained a slightly higher mean score than the Desktop condition, although the difference was small.

Table 5. SUS scores for the Desktop and HMD-based conditions ($n = 36$).

Statistic	Desktop condition	HMD-based condition
Mean	78.19	79.72
SD	11.44	8.80
Median	76.25	80.00
IQR	70.00–87.50	72.50–85.63

The mean paired difference was $+1.53$ points in favor of the HMD-based condition. Among the 36 participants, 18 gave a higher SUS score to the HMD-based condition, 12 gave a higher score to the Desktop condition, and 6 gave equal scores to both conditions. The Wilcoxon signed-rank test did not indicate a statistically significant difference between conditions ($p = .503$, $r_{rb} = .140$). Thus, the HMD-based condition preserved perceived usability at a level comparable to the Desktop condition.

4.3 User Experience

Table 6 presents the UEQ scale scores for both conditions. All UEQ scale means were positive in both conditions, indicating favorable user experience overall. The HMD-based condition obtained higher mean values for Attractiveness, Perspicuity, Stimulation, and Novelty, whereas the Desktop condition obtained higher values for Efficiency and Dependability.

The largest descriptive differences favoring the HMD-based condition occurred in Novelty and Stimulation, whereas the largest descriptive difference favoring the Desktop condition occurred in Efficiency. However, none of the UEQ scale differences remained statistically significant after Holm–Bonferroni correction. Therefore, these results should not be interpreted as statistically confirmed superiority of either condition on UEQ dimensions. Instead, they indicate a descriptive pattern in which the HMD-based condition tended toward stronger hedonic ratings and the Desktop condition tended toward higher perceived efficiency.

4.4 Cybersickness

The SSQ was applied after the HMD-based condition. Table 7 presents the SSQ subscale and total scores. The mean total SSQ score was 32.41, with a median of 22.44. Three participants reported no symptoms. The highest subscale scores were observed for Disorientation, followed by Oculomotor symptoms and Nausea.

Three participants reported no symptoms. The most frequent symptoms were fullness of head, reported by 21 participants; eyestrain, reported by 20 participants; dizziness with eyes open, reported by 20 participants; blurred vision, reported by 18 participants; and general discomfort, reported by 16 participants. These results indicate that the HMD-based condition was strongly preferred, but not uniformly comfortable. The mean total SSQ score, the high Disorientation subscale, and the frequency of symptoms such as fullness of head, eyestrain, dizziness with eyes open, blurred vision, and general discomfort show that cybersickness and visual/vestibular

Table 6. UEQ scale scores for the Desktop and HMD-based conditions ($n = 36$).

UEQ scale	Desktop Mean	Desktop SD	HMD Mean	HMD SD	Δ HMD–Desktop	p_{Holm}
Attractiveness	1.41	0.66	1.54	0.48	+0.13	0.779
Perspicuity	2.14	0.93	2.15	0.84	+0.01	1.000
Efficiency	2.31	0.64	2.07	0.61	-0.24	0.360
Dependability	1.88	0.78	1.83	0.79	-0.06	1.000
Stimulation	1.77	1.02	2.01	0.86	+0.24	0.693
Novelty	1.02	1.22	1.31	0.96	+0.29	0.258

Table 7. SSQ scores after the HMD-based condition ($n = 36$).

SSQ score	Mean	SD	Median	IQR
Nausea	15.90	16.13	9.54	0.00–28.62
Oculomotor	24.64	22.00	18.95	7.58–32.22
Disorientation	52.20	49.17	27.84	13.92–83.52
Total	32.41	28.12	22.44	14.96–48.62

discomfort were relevant aspects of the HMD-based experience. Therefore, preference for the HMD-based condition should be interpreted together with comfort limitations rather than as evidence that the immersive condition was unproblematically accepted.

4.5 Objective Task Performance

Objective task performance was computed from the paired telemetry dataset with 36 matched participants. The logs provide stage-level values for each board, including duration, attempts, hits, and errors; these records were used to reconstruct total completion time, total attempts, total hits, total errors, precision, and hesitation index. The use of stage-level records is particularly important because the HMD telemetry file stores valid stage data even when final aggregate fields are zero, and the Desktop file also contains records where stage-level values are more reliable than final summary fields.

Table 8 presents the global objective metrics. The Desktop condition was faster on average and in median completion time. Participants completed the task in 354.00 seconds on average in the Desktop condition and 436.97 seconds in the HMD-based condition, corresponding to a mean difference of +82.97 seconds for HMD-based interaction. However, this difference did not remain statistically significant after Holm–Bonferroni correction across telemetry metrics ($p_{Holm} = .119$). Attempts, errors, precision, and hesitation index also showed no reliable corrected differences. Therefore, the objective results should be interpreted as evidence of a descriptive time-cost tendency for HMD-based interaction, not as a statistically confirmed global performance penalty.

For attempts, errors, precision, and hesitation, the differences were small and not statistically significant. The HMD-based condition required slightly more attempts and produced a slightly higher hesitation index, while the Desktop condition showed slightly higher precision. However, these differences did not approach significance after correction. Overall, the objective data suggest that the clearest performance difference was time, whereas accuracy-related measures were broadly comparable between conditions.

4.6 Stage-Level Performance

To identify whether the performance differences were distributed across the entire task or concentrated in specific stages, completion time was also analyzed by map board. Table 9 presents the stage-level completion times.

The largest descriptive time difference occurred in the North board, where the HMD-based condition required 126.57 seconds on average compared with 84.27 seconds in the Desktop condition. However, this difference did not remain statistically significant after Holm correction ($p_{Holm} = .309$). The Northeast board also showed a descriptive increase in HMD completion time, but the corrected result was not significant. Therefore, the stage-level results suggest that the HMD time cost was more visible in some boards, especially North and Northeast, but no stage-level difference was statistically reliable after correction.

4.7 Preference and Qualitative Feedback

The post-experience questionnaire showed a clear preference for the HMD-based condition. Overall, 32 of the 36 participants preferred the HMD-based condition, corresponding to 88.89% of the sample, whereas 4 participants preferred the Desktop condition. An exact binomial test against an equal-preference baseline indicated that this preference distribution was statistically significant ($p < .001$). Thus, preference for the HMD-based condition was one of the strongest findings of the study.

Open-ended responses help explain this result. Several participants described the HMD-based condition as fun, innovative, creative, or suitable for use in schools. Other comments suggested design improvements, including clearer feedback when a piece is incorrectly placed, showing the names of states or regions, separating pieces more clearly before placement, increasing the size of small target areas such as the Distrito Federal, and fixing occasional interaction bugs. These comments indicate that the HMD-based condition was strongly preferred for its novelty and engagement, but that participants still noticed concrete usability and feedback issues.

4.8 Subjective–Objective Associations

Exploratory Spearman correlations were computed to examine whether subjective ratings were associated with objective performance in the paired telemetry sample. No strong or statistically significant associations were observed. In the HMD-based condition, SUS was not meaningfully correlated with completion time ($\rho = -.011$, $p = .952$) or precision ($\rho = .059$, $p = .751$). UEQ Efficiency was also not correlated with HMD completion time ($\rho = .069$, $p = .713$).

Table 8. Objective performance metrics reconstructed from stage-level telemetry ($n = 36$ paired participants).

Metric	Desktop Mean	Desktop Median	HMD Mean	HMD Median	Δ HMD–Desktop	p_{Holm}
Total time (s)	354.00	343.50	436.97	397.23	+82.97	0.119
Attempts	45.36	40.50	48.56	42.50	+3.19	1.000
Errors	10.17	7.00	10.11	6.50	-0.06	1.000
Precision	0.742	0.790	0.717	0.753	-0.025	1.000
Hesitation index	13.36	8.50	16.56	10.50	+3.19	1.000

Table 9. Stage-level completion time in seconds for the Desktop and HMD-based conditions ($n = 36$).

Stage	Desktop Mean	Desktop Median	HMD Mean	HMD Median	Δ HMD–Desktop	p	p_{Holm}
Macro-regions	61.50	52.92	64.37	47.86	+2.87	0.798	1.000
North	84.27	63.55	126.57	87.95	+42.29	0.052	0.309
Northeast	121.46	88.52	140.95	111.19	+19.49	0.123	0.614
Midwest	48.37	40.13	58.89	50.85	+10.52	0.614	1.000
Southeast	21.17	18.73	26.23	20.32	+5.06	0.375	1.000
South	17.23	14.97	19.97	15.48	+2.73	0.550	1.000

Table 10. Post-experience simulator preference ($n = 36$).

Preferred condition	Participants	Percentage
HMD-based condition	32	88.89%
Desktop condition	4	11.11%

Similarly, SSQ total score did not correlate significantly with HMD completion time ($\rho = -.073$, $p = .697$) or hesitation index ($\rho = .147$, $p = .429$).

These results suggest that preference for the HMD-based condition was not simply a reflection of faster or more accurate performance. Instead, the overall pattern indicates a trade-off: the HMD-based condition was strongly preferred and was rated as comparably usable, but the telemetry suggests that it often required more time to complete the same task. This supports the interpretation that the HMD-based condition provided a hedonic and engagement advantage, while the Desktop condition retained descriptive advantages in interaction efficiency and lower adaptation demand.

4.9 Effect of Condition Order

Because participants completed the same cartographic task in both conditions, we analyzed whether presentation order influenced the results. The updated dataset was balanced, with 18 participants starting with the HMD-based condition and 18 starting with the Desktop condition.

Order influenced completion time. In the HMD-first group, the HMD-based condition was slower than the Desktop condition by 165.22 seconds on average and 157.67 seconds in the median. In contrast, in the Desktop-first group, the mean paired difference between HMD and Desktop was only 0.72 seconds, with a median difference of 2.47 seconds. A Mann–Whitney test on the paired HMD–Desktop time differences indicated that the time difference varied by order group ($p = .020$, Cliff's $\delta = .457$).

A complementary first-versus-second analysis showed that the second condition was faster than the first condition by 82.25 seconds on average and 87.84 seconds in the median. The Wilcoxon signed-rank test indicated that this first-to-second reduction was statistically significant ($p = .028$, $r_{rb} = -.417$). This result suggests a learning or carryover effect: participants became faster after completing the task once, regardless of the interface used first.

Order did not meaningfully change post-experience preference. In the HMD-first group, 17 of 18 participants preferred the HMD-based condition, while in the Desktop-first group, 15 of 18 participants preferred the HMD-based condition. Fisher's exact test did not indicate a significant association between order and preference ($p = .603$). Therefore, preference for the HMD-based condition remained robust, whereas completion-time comparisons were sensitive to condition order.

Exploratory analyses did not indicate reliable order effects for attempts, errors, precision, or hesitation. Subjective outcomes were also less clearly affected by order. SUS differences did not vary significantly by order group ($p = .390$). UEQ Perspicuity and Efficiency showed exploratory order-related tendencies, with the HMD-based condition receiving relatively lower pragmatic scores when it was experienced first, but these patterns should be interpreted cautiously because they were secondary analyses. SSQ scores were descriptively higher when the HMD-based condition was experienced second, but this result was also exploratory and should not be treated as conclusive.

5 Discussion

This study compared the Desktop and HMD version conditions of GeoConnect using subjective questionnaires, post-experience preference, and telemetry-based performance metrics. Overall, the results indicate that the HMD-based condition was strongly preferred and perceived as comparably usable to the Desktop condition, but this preference did not translate into faster objective performance. Instead, the telemetry suggests a trade-off: immersive interaction increased engagement and preference while imposing additional time cost in the same cartographic task.

5.1 Usability Was Preserved Across Interaction Conditions

The first research question asked whether the two conditions differed in perceived usability. Both conditions obtained positive SUS scores, and the difference between them was small and not statistically significant. This result suggests that the HMD-based condition did not introduce a substantial usability penalty, even though it required participants to use a less conventional interaction modality based on tracked controllers and spatial manipulation.

This finding is relevant because HMD-based VR systems often introduce additional interaction demands, including headset adjustment, controller use, spatial movement, and adaptation to stereoscopic visualization. In this study, these factors did not prevent participants from evaluating the HMD-based condition as usable. One possible explanation is that the GeoConnect task is based on a simple and familiar interaction metaphor: selecting, moving, and placing pieces in a target area. Although this metaphor is implemented differently in the Desktop and HMD-based conditions, its conceptual structure remains understandable across both conditions.

At the same time, the similarity in SUS scores should not be interpreted as evidence that the two interfaces were experienced in the same way. SUS captures a global perception of usability, but it does not fully separate efficiency, confidence, enjoyment, novelty, and physical comfort. For this reason, the UEQ and telemetry results provide a more nuanced interpretation of the interaction trade-offs.

5.2 A Pragmatic–Hedonic Trade-Off Between the Desktop and HMD-Based Conditions

The second research question examined user experience across pragmatic and hedonic dimensions. The UEQ results showed positive scores for both conditions. Descriptively, the HMD-based condition obtained higher values for Attractiveness, Stimulation, and Novelty, whereas the Desktop condition obtained higher values for Efficiency and Dependability. Although these differences did not remain statistically significant after correction, the pattern is theoretically meaningful and consistent with previous educational VR research.

Immersive VR is often associated with presence, engagement, and affective value, but these qualities do not automatically imply better task efficiency or learning outcomes [Cummings and Bailenson, 2016; Parong and Mayer, 2018; Makransky and Petersen, 2021]. The present results follow this pattern. Participants tended to perceive the HMD version as more stimulating and novel, which helps explain the strong preference for this condition. However, the Desktop version retained a descriptive advantage in efficiency, possibly because mouse-based interaction was stable and required less physical adaptation in this task.

This distinction is important for educational application. If the goal is to maximize engagement, curiosity, and willingness to participate, HMD-based VR may offer clear advantages. If the goal is to maximize speed and procedural fluency, a Desktop interface may remain competitive or even preferable. Therefore, the choice between Desktop and HMD-based VR should depend on the intended educational use case rather than on an assumption that immersion is always superior.

5.3 Objective Performance Suggests a Time-Cost Tendency for the HMD-Based Condition

The third research question focused on objective task performance. The clearest telemetry pattern was a descriptive tendency for the HMD-based condition to require more time than the Desktop condition. The mean completion time was 82.97 seconds higher in the HMD-based condition, but this difference did not remain statistically significant after Holm–Bonferroni correction across telemetry metrics. Therefore, the result should be interpreted as a time-cost tendency moderated by order and familiarization, not as definitive evidence that HMD-based interaction is inherently slower.

This pattern suggests that the HMD-based condition did not substantially reduce correctness, but it increased the time needed to complete the task. In other words, participants were generally able to solve the same map-assembly activity in immersive VR, but they did so more slowly. This additional time may reflect several factors: adaptation to 6-DoF controllers, physical movement required to inspect or manipulate pieces, slower spatial positioning, controller tracking limitations, or the need to visually search the environment from within the headset.

The stage-level analysis reinforces this interpretation. The largest time difference occurred in the North board, followed descriptively by the Northeast board. These stages may have imposed higher visual search or manipulation demands, making the cost of immersive interaction more visible. The effect was not uniform across all boards, which suggests that the HMD-based condition was not globally worse, but more sensitive to the characteristics of specific stages. Boards with more pieces, smaller targets, visually similar shapes, or less familiar geographic content may amplify the difference between mouse-based and controller-based manipulation.

An important implication is that objective performance should not be reduced to completion time alone. The HMD-based condition was slower on average, but not clearly less accurate. This distinction matters for educational applications: spending more time on a task can indicate inefficiency, but it can also reflect deeper engagement, exploration, or more deliberate interaction. Without a direct learning assessment, such as a pre/post cartography test, it is not possible to determine whether the additional time in HMD was educationally beneficial or merely an interaction cost.

The time increase observed in the HMD-based condition may also be related to comfort constraints. Although the exploratory correlations did not show a statistically significant association between SSQ total score and HMD completion time, the SSQ results indicate non-negligible visual and vestibular discomfort. Symptoms such as dizziness, blurred vision, eyestrain, and fullness of head could plausibly slow visual inspection, locomotion, or object manipulation for some participants. Therefore, cybersickness should be considered a possible contributor to interaction cost, even though the present study does not establish it as a confirmed mediator.

5.4 Order Effects and Familiarization

The order analysis suggests that the apparent time cost of HMD-based interaction was strongly influenced by familiarization. Participants who experienced the HMD-based condition first were substantially slower in HMD than in Desktop, whereas participants who experienced Desktop first completed the later HMD-based condition with nearly equivalent times. This pattern suggests that the HMD-based interface may impose an initial adaptation cost related to controller use, embodied manipulation, locomotion, or novelty. At the same time, the first-to-second reduction across conditions indicates that participants became faster after completing the task once, independently of the interface order.

This finding has direct practical implications. HMD-based educational activities should include a short familiarization phase before the main learning task, especially when students are new to 6-DoF controllers or artificial locomotion. A brief practice task may reduce the interaction cost and help separate learning about the educational content from learning how to operate the interface.

5.5 Preference for HMD Was Strong Despite Slower Performance

The fifth research question examined preference and qualitative feedback. The post-experience questionnaire showed a clear preference for the HMD-based condition, with 32 of 36 participants choosing it over the Desktop condition. This is one of the strongest findings of the study. Importantly, preference was not meaningfully affected by presentation order: 17 of 18 participants in the HMD-first group and 15 of 18 participants in the Desktop-first group preferred the HMD version. This indicates that the preference for immersive VR was robust even though objective completion time was sensitive to order.

This result highlights a common tension in VR evaluation: subjective preference may be driven by engagement, novelty, embodiment, and enjoyment rather than by efficiency alone. For educational application, this tension is not necessarily negative. Motivation and willingness to interact can be valuable, especially in school contexts where student engagement is a central challenge. However, preference should not be treated as sufficient evidence of pedagogical superiority. A game can be preferred because it is more fun or novel, while still requiring design improvements to support learning, efficiency, and accessibility. However, in geography education, immersive digital environments can promote exploratory and experimental engagement with geographic concepts, even they impose additional interactions demands.

The open-ended comments support this interpretation. Participants frequently described the HMD-based condition as fun, innovative, and suitable for school use. At the same time, they suggested practical improvements, such as clearer feedback for incorrect placements, larger or more visible elements, labels for states or regions, and corrections to occasional interaction problems. These comments show that preference for HMD coexisted with awareness of usability limitations.

5.6 Cybersickness and Comfort Should Be Treated as Design Constraints

The fourth research question addressed cybersickness after HMD use. The SSQ results indicated that symptoms were present for a non-negligible portion of the sample, especially disorientation, oculomotor symptoms, eyestrain, dizziness, blurred vision, and fullness of head. Although these symptoms did not prevent strong preference for the HMD-based condition, they remain important for classroom adoption.

This is especially relevant because the participants were young students. In school settings, VR activities must be short, supervised, and designed to minimize discomfort. Even moderate symptoms can affect willingness to reuse the system, concentration during the activity, and the feasibility of deploying the application with larger groups. Therefore, the HMD version condition should not be evaluated only by usability and preference, but also by comfort and safety.

Several design choices may help reduce cybersickness and discomfort in future versions of GeoConnect. The task should minimize unnecessary locomotion, keep the main interaction area stable and visible, avoid rapid camera movement, provide sufficient tutorial time, and allow users to complete the task while standing or seated comfortably. Since the main activity is map assembly rather than navigation, locomotion should support inspection without becoming a source of disorientation.

5.7 Implications for Educational VR Design

The results lead to several design implications for educational VR cartography application involving spatial manipulation, listed in Table 11.

5.8 Implications for Cartography Learning

GeoConnect was designed around active map assembly, which requires learners to identify spatial relationships between Brazilian macro-regions and states. The strong preference for the HMD-based condition suggests that immersive VR can make this type of cartographic activity more attractive to students. This is valuable because motivation and engagement may increase opportunities for practice and discussion in the classroom.

However, the present study did not directly measure learning gains. Therefore, the results support claims about usability, user experience, preference, comfort, and interaction performance, but not yet about improved cartographic knowledge. Future studies should include pre- and post-tests assessing recognition of regions, state locations, acronyms, and spatial relationships. Such assessments would make it possible to determine whether the additional time and engagement observed in HMD translate into better learning outcomes.

A promising direction is to evaluate whether immersive VR benefits specific learning objectives more than others. For example, HMD-based interaction may be especially useful for spatial orientation, geographic organization, and collaborative exploration, while Desktop interaction may be sufficient for memorization or quick practice. Rather than asking whether VR is globally better than Desktop, future

Table 11. Design implications for educational cartography applications using Desktop and HMD-based interaction.

Finding	Design implication	Practical recommendation
Most participants preferred the HMD-based condition.	Immersion can support motivation, attention, and perceived engagement.	Use HMD-based activities when novelty, spatial engagement, and exploratory motivation are central goals.
The HMD-based condition was slower on average, but the difference did not survive Holm correction.	Immersive interaction may introduce an initial time cost, but the evidence should be interpreted cautiously.	Allocate extra time for HMD activities and avoid treating speed as the only success criterion.
Completion time depended strongly on presentation order.	Familiarization and task-learning effects influence HMD performance.	Include a short practice activity before the main task, especially for students with limited HMD experience.
Both conditions achieved high SUS scores.	Desktop and HMD-based conditions can both be usable for the same educational task.	Choose the modality according to learning goals, available time, equipment, classroom logistics, and comfort constraints.
SSQ symptoms were non-negligible in the HMD-based condition.	Comfort is a design constraint, not a secondary detail.	Keep sessions short, minimize unnecessary continuous locomotion, monitor discomfort, and provide breaks or alternative Desktop access.
No direct pre/post cartography test was applied.	Interaction outcomes cannot be equated with learning gains.	Future studies should include direct learning measures to determine whether additional HMD time reflects educational benefit or interaction cost.

work should identify which educational objectives benefit most from immersive interaction.

5.9 Threats to Validity

Some threats to validity should be considered. First, the comparison between Desktop and HMD-based VR does not isolate stereoscopy alone. The HMD-based condition also changes field of view, input modality, head tracking, embodiment, locomotion, and visual inspection strategy. Therefore, the results should be interpreted as a comparison between two complete interaction configurations rather than as a pure test of stereoscopic visualization.

Second, condition order influenced completion time. Although the experiment counterbalanced the starting condition, participants completed the same cartographic task twice. The updated order analysis showed that the second condition was faster overall, indicating a learning or carryover effect. Therefore, completion-time differences should not be interpreted as a pure interface effect. Future studies should use alternate but equivalent task versions, randomized stage orders, or additional training trials to reduce repeated-task learning.

Third, the sample consisted of students from a specific age range and school context. The findings may not generalize to younger children, older students, adults, or users with different levels of geographic knowledge or VR familiarity. Prior HMD experience may also influence adaptation to immersive interaction, although the present sample was not large enough to support detailed subgroup analysis.

Fourth, the study relied on self-report instruments for usability, user experience, and cybersickness. Although SUS, UEQ, and SSQ are standardized and widely used, they remain subjective measures and may be influenced by novelty, social desirability, and participant expectations.

Fifth, the objective analysis required cleaning and matching telemetry sessions across data sources. In the updated dataset, all 36 participants had complete paired telemetry records, and the previous implausible duration value was corrected. Nevertheless, the analysis still depended on accurate stage-level logging and reconstruction of global metrics from telemetry records. Future versions should include automated validation checks for missing stages, duplicated sessions, zero durations, and implausible timestamps.

A further limitation is that the study did not include a direct pre/post assessment of cartography learning. Consequently, the additional time observed descriptively in the HMD-based condition cannot be interpreted as educationally beneficial time-on-task. It may reflect deeper spatial engagement, but it may also reflect controller adaptation, artificial locomotion, visual search difficulty, cybersickness, or general interaction cost. Future work should combine the present subjective and telemetry measures with direct learning outcomes, such as pre/post cartography tests, delayed retention, or transfer tasks.

The study measured prior HMD exposure but did not separately quantify everyday desktop/laptop, mouse/keyboard, or mobile-device use. Therefore, claims about Desktop familiarity should be treated cautiously, especially because students in this age range may be mobile-first users.

5.10 Summary of Findings

Taken together, the results suggest that the HMD-based condition is promising but not unconditionally superior for GeoConnect. The HMD-based condition was strongly preferred and preserved perceived usability, indicating that participants accepted and enjoyed the headset-based experience. At the same time, telemetry suggested that the same task often required more time in HMD, although this difference was moderated by condition order and did not remain significant after Holm correction across telemetry metrics. Therefore, the performance results should be interpreted cautiously as evidence of an interaction-and-learning trade-off rather than as a pure interface effect.

Therefore, the main contribution of this comparison is not to declare one condition universally better than the other, but to clarify their trade-offs. The Desktop condition showed descriptive advantages in efficiency and logistical accessibility, while the HMD-based condition offered stronger engagement, novelty, and experiential appeal. For educational deployment, the best use of GeoConnect may involve combining both modalities: the Desktop condition for quick, accessible practice and the HMD-based condition for immersive, motivating, and exploratory learning sessions.

6 Conclusion and Future Work

The results show that both conditions were perceived positively in terms of usability. The HMD-based condition achieved a SUS score comparable to the Desktop condition, indicating that immersive interaction did not introduce a substantial perceived usability penalty. UEQ results further suggested a pragmatic–hedonic trade-off: the HMD-based condition was descriptively stronger in stimulation, novelty, and attractiveness, while the Desktop condition showed descriptive advantages in efficiency and dependability. This pattern reinforces the idea that immersive VR should not be evaluated only as a more advanced display technology, but as an interaction configuration that changes engagement, control, effort, and task execution.

The telemetry results indicate that the HMD-based condition was slower on average than the Desktop condition, with a mean difference of +82.97 seconds. However, this time-cost tendency did not remain statistically significant after Holm–Bonferroni correction across telemetry metrics ($p_{Holm} = .119$), and it was strongly moderated by presentation order. Therefore, the results should not be read as definitive evidence that HMD-based interaction is inherently slower. Instead, they suggest that immersive interaction may introduce an initial familiarization cost that can be reduced when students have already learned the task structure. Attempts, errors, precision, and hesitation index also did not show reliable corrected differences.

Cybersickness results showed that symptoms after HMD use were present for part of the sample, especially oculomotor and disorientation-related symptoms. Although these symptoms did not prevent strong preference for the HMD-based condition, they should be treated as a central design constraint for future classroom use. Educational VR applications for young users should minimize unnecessary locomotion, provide stable visual references, offer clear interaction feedback, and keep activities within appropriate exposure durations.

Overall, the findings support a complementary interpretation: the Desktop condition is suitable when time efficiency, accessibility, and logistical simplicity are priorities, whereas the HMD-based condition may be valuable when engagement, novelty, and spatial exploration are central instructional goals. Future work should include direct cartography learning assessments and more detailed characterization of participants' prior desktop, mobile, and HMD experience.

Future work should extend this evaluation in several directions. First, future studies should include direct learning measures, such as pre- and post-tests on Brazilian regions, state locations, acronyms, and spatial relationships. This would allow researchers to determine whether the engagement and additional task time observed in the HMD-based condition translate into measurable learning gains. Second, future versions of GeoConnect should improve interaction feedback, including clearer visual, auditory, or haptic cues for correct and incorrect placements. Third, adaptive scaffolding mechanisms should be explored, such as optional labels, hints, progressive difficulty, magnification of small regions, and assistance for visually similar states.

Further studies should also investigate the role of previous VR experience, condition order, and individual differences in spatial ability. Larger samples would allow subgroup analyses comparing students with and without previous HMD experience, as well as potential gender or age-related effects. In addition, future work should examine alternative locomotion and manipulation techniques to reduce completion time and cybersickness while preserving the motivational benefits of immersive VR. Finally, classroom-based longitudinal studies are needed to understand how GeoConnect performs when integrated into real teaching practice, including teacher mediation, group dynamics, repeated use, and curricular alignment. Although the present study evaluated usability, user experience, and task performance, it did not directly measure learning gains related to cartographic literacy or spatial reasoning. Future studies should investigate whether immersive interaction contributes to measurable improvements in spatial-thinking development and cartographic learning outcomes through pre-test/post-test education protocols.

Declarations

Author contributions

LZO: Writing - original draft, Conceptualization, Formal analysis. **MPS, GCD, TNJ**: Writing - review. **CRMM, EHH, JMT, FLSN, FFFP**: Supervision, Writing - review & editing. All authors read and approved the final manuscript.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We would like to thank everyone who contributed to the data collection process of this study, particularly Luis Fernando Cavalcante da Silva and Ana Carolina Verona. The participation, availability, and collaboration of all those involved were essential to the development of this research. Without their support, it would not have been possible to carry out this important stage of the study and achieve the results presented. We express our sincere appreciation and gratitude to everyone who, in some way, contributed to making this research possible.

Funding

The authors have no funding to report for this research.

Availability of data and materials

Data and materials supporting this study are available upon request.

Use of Artificial Intelligence

Generative Artificial Intelligence tools were used to support the initial reorganization of the manuscript and the revision of the writing, always under full human supervision. Decisions regarding scope, method, interpretation of results, final wording, and responsibility for the content remained exclusively with the authors.

References

Bimberg, P., Weissker, T., and Kulik, A. (2020). On the usage of the simulator sickness questionnaire for virtual reality research. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 464–467. IEEE. DOI: 10.1109/VRW50115.2020.00098.

Brooke, J. *et al*. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7. Available at: `https://www.researchgate.net/publication/319394819_SUS_--_a_quick_and_dirty_usability_scale`.

Carbonell-Carrera, C. and Saorin, J. L. (2017). Virtual learning environments to enhance spatial orientation. *Eurasia Journal of Mathematics, Science and Technology Education*, 14(3):709–719. DOI: 10.12973/ejmste/79171.

Chen, M.-X., Hu, H., Yao, R., Qiu, L., and Li, D. (2024). A survey on the design of virtual reality interaction interfaces. *Sensors*, 24(19):6204. DOI: 10.3390/s24196204.

Cummings, J. J. and Bailenson, J. N. (2016). How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. *Media psychology*, 19(2):272–309. DOI: 10.1080/15213269.2015.1015740.

Czimre, K., Teperics, K., Molnár, E., Kapusi, J., Saidi, I., Gusman, D., and Bujdosó, G. (2024). Potentials in using vr for facilitating geography teaching in classrooms: a systematic review. *ISPRS International Journal of Geo-Information*, 13(9):332. DOI: 10.3390/ijgi13090332.

Dede, C. (2009). Immersive interfaces for engagement and learning. *science*, 323(5910):66–69. DOI: 10.1126/science.1167311.

Edler, D., Wiktorin, D., Becker, G., Schweins, U., Reichert, K., Keil, J., and Dickmann, F. (2025). Potential of immersive virtual reality for teaching geography in schools. *KN-Journal of Cartography and Geographic Information*, pages 1–13. DOI: 10.1007/s42489-025-00204-1.

Harms, P. (2019). Automated usability evaluation of virtual reality applications. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(3):1–36. DOI: 10.1145/3301423.

Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool. DOI: 10.1145/2792790.

Kablitz, D. (2025). Bridging theory and practice with immersive virtual reality: A study on transfer facilitation in vet. *Education Sciences*, 15(8):959. DOI: 10.3390/educsci15080959.

Kamińska, D., Zwoliński, G., and Laska-Leśniewicz, A. (2022). Usability testing of virtual reality applications—the pilot study. *Sensors*, 22(4):1342. DOI: 10.3390/s22041342.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220. DOI: 10.1207/s15327108ijap0303$_3$.

Laugwitz, B., Held, T., and Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*, pages 63–76. Springer. DOI: 10.1007/978-3-540-89350-9$_6$.

Liu, C., Meng, S., Zheng, W., and Zhou, Z. (2025). Research on the impact of immersive virtual reality classroom on student experience and concentration. *Virtual Reality*, 29(2):82. DOI: 10.1007/s10055-025-01153-w.

Makransky, G. and Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, 66(5):1141–1164. DOI: 10.1007/s11423-018-9581-2.

Makransky, G. and Petersen, G. B. (2019). Investigating the process of learning with desktop virtual reality: A structural equation modeling approach. *Computers & Education*, 134:15–30. DOI: 10.1016/j.compedu.2019.02.002.

Makransky, G. and Petersen, G. B. (2021). The cognitive affective model of immersive learning (camil): A theoretical research-based model of learning in immersive virtual reality. *Educational psychology review*, 33(3):937–958. DOI: 10.1007/s10648-020-09586-2.

Mayer, R. E. (2019). Computer games in education. *Annual review of psychology*, 70(1):531–549. DOI: 10.1146/annurev-psych-010418-102744.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., and Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in k-12 and higher education: A meta-analysis. *Computers & education*, 70:29–40. DOI: 10.1016/j.compedu.2013.07.033.

Oliveira, L. d. (2011). Estudo metodológico e cognitivo do mapa. In Almeida, R. D. d., editor, *Cartografia Escolar*. Contexto, São Paulo. Book.

Parong, J. and Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of educational psychology*, 110(6):785. DOI: 10.1037/edu0000241.

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

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

Shelton, B. E. and Hedley, N. R. (2002). Using augmented reality for teaching earth-sun relationships to undergraduate geography students. In *The First IEEE International Workshop Agumented Reality Toolkit,*, pages 8–pp. IEEE. DOI: 10.1109/ART.2002.1106948.

Simielli, M. E. R. (2011). O mapa como meio de comunicação e a alfabetização cartográfica. In Almeida, R. D. d., editor, *Cartografia Escolar*. Contexto, São Paulo. Book.

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

Souchet, A. D., Lourdeaux, D., Pagani, A., and Rebenitsch, L. (2023). A narrative review of immersive virtual reality's ergonomics and risks at the workplace: cybersickness, visual fatigue, muscular fatigue, acute stress, and mental overload. *Virtual Reality*, 27(1):19–50. DOI: 10.1007/s10055-022-00672-0.

Stanney, K. M., Kennedy, R. S., and Drexler, J. M. (1997). Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, volume 41, pages 1138–1142. SAGE Publications Sage CA: Los Angeles, CA. DOI: 10.1177/107118139704100292.

Střechová, M., Černỳ, M., Šašinka, Č., Stachoň, Z., Šašinková, A., Holubec, F., and Švédová, H. (2025). Collaborative immersive virtual environments in geography education on climate zones: A ux case study. *ISPRS International Journal of Geo-Information*, 14(12):455. DOI: 10.3390/ijgi14120455.