ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.2730

Virtual Reality applied to teaching the assembly of prefabricated housing in graduate studies: a mixed analysis of usability and performance

José Requesens[1]; Reinaldo Valdebenito[2]; Francisco Muñoz[3]; Juan C. Gallegos[4] Eric Forcael[5]
[1,3]Pontificia Universidad Católica de Valparaíso, Chile, *jose.requesens@pucv.cl, francisco.munoz@pucv.cl*
[2]Universidad Técnica Federico Santa María, Chile, *reinaldo.valdebenito@usm.cl,*
[4,5]Universidad San Sebastián, Chile, *jgallegosp3@correo.uss.cl, eric.forcael@uss.cl*

Abstract*– Teaching the assembly of prefabricated housing requires mastering both procedural sequencing and spatial understanding to position and connect structural components accurately. This paper evaluates a guided immersive Virtual Reality (VR) workflow designed to simulate the structural assembly of a prefabricated house as a safe, repeatable learning activity. In an exploratory single-session pilot using a criterion-based purposive sample of 12 graduate students, participants completed the VR assembly training task and responded to a post-experience questionnaire (Likert 1-7) along with open-ended questions, while the VR software automatically logged objective performance indicators (time, attempts, successes/errors, accuracy, and efficiency). Survey results showed consistently high ratings across usability/conditions, perceived learning, and transfer intention (an overall rating of 6.25 out of 7). Software logs indicated an average completion time of 10.69 ± 2.65 minutes and a mean accuracy of 0.61 ± 0.18, with significant variability in error-driven rework (ranging from 3 to 40), suggesting diverse execution strategies. Correlation analyses confirmed strong internal links among objective metrics (e.g., accuracy vs. errors/min ρ = -0.984), while perception-performance associations were generally weak; notably, perceived learning increased with time/attempts and decreased with accuracy/efficiency, suggesting a "productive struggle" pattern during early exposure. Qualitative feedback emphasized improvements in room-scale safety, practice recurrence, and task complexity. Overall, the study provides preliminary evidence that a triangulated evaluation framework can support the assessment of immersive VR training beyond satisfaction-based measures, linking perceived experience with measurable procedural performance and practical design improvements.

***Keywords*-- Virtual Reality, Construction Education, Prefabrication, Usability, Performance Metrics.**

I. Introduction

Teaching the assembly of prefabricated housing requires both procedural steps and spatial understanding to accurately position and connect structural components. However, students often struggle to translate conceptual explanations into practical spatial reasoning during construction. In graduate programs, where an immediate application to real construction scenarios is expected, this gap often persists. When instruction is limited to traditional formats such as lectures, floor plans, or videos, students may understand the process conceptually but struggle to grasp the assembly, anticipate interferences, and recognize common alignment or anchoring errors. VR is especially relevant for this type of task because it can create a clear illusion of 'being there' in the scenario where the problem is encountered, thereby promoting more realistic responses in real-time [1].

In this context, VR has established itself as a safe, controlled, and repeatable training alternative, with clear advantages for practicing complex procedures without exposing individuals to risks or relying on physical prototypes [2]. Additionally, its standardization facilitates scalability, providing the same experience to all students and allowing for the capture of objective evidence of performance during the task. In fact, systematic reviews and meta-analyses in education and training report overall positive effects of VR programs, although they depend heavily on design and the 'fit' between the task and the technology [3], [4].

However, evidence also suggests that immersion alone does not guarantee better learning outcomes. Meta-analyses of immersive VR show moderate or small average effects, with variation depending on educational level and implementation type [5]. Experimental studies have shown that a greater sense of presence can coexist with higher cognitive load and lower learning outcomes when the instructional design does not control distractions or overload [6]. In line with this, recent models explain how features such as immersion and interactivity influence presence, agency, and cognitive-affective processes that can either facilitate or hinder learning [7], [8].

This work presents a VR experience aimed at teaching the assembly of a prefabricated house, applied to final-semester students in the master's in construction management and sustainability at the San Sebastián University (USS) in Chile. The study adopts a mixed approach based on: (i) a student perception questionnaire that evaluates the quality of immersive experience and its ability to generate learning; and (ii) objective metrics recorded by the software during the activity, such as time, attempts, successes, errors, accuracy, and efficiency. This study is intentionally framed as an exploratory pilot. Rather than seeking statistical generalization or causal inference, it aims to assess the feasibility and analytical value of integrating self-reported perceptions, software-recorded performance indicators, and qualitative feedback within an authentic postgraduate immersive VR learning context.

In particular, the study is guided by the following research questions: a) RQ1: How do graduate students evaluate the usability, immersion, and learning conditions of the VR experience?; b) RQ2: What level of perceived learning and transfer potential to real-world contexts do students report after the experience?, and c) RQ3: What is the relationship between the perceptions reported regarding the experience and learning, and the objective performance recorded by the software?

II. State of the Art

This section reviews the main research streams that frame this study: Construction 4.0 and its increasing demand for practice-oriented digital skills; the use of VR in Architecture, Engineering, and Construction (AEC) for both collaboration and training; and evidence on the effectiveness of immersive learning in higher education. The review then positions this work's contribution by identifying the specific gap it addresses: graduate VR training for prefabricated housing assembly, evaluated using a mixed-methods approach that combines perceived experience and learning with software-recorded performance metrics.

A. Construction 4.0 and skills requirements

Construction 4.0 describes the ongoing integration of digital technologies, such as Building Information Modeling (BIM), the Internet of Things (IoT), robotics, Artificial Intelligence (AI), and cyber-physical systems, throughout the construction lifecycle to enhance productivity, quality, cost management, and sustainability. In developing countries, the concept remains diverse, and its adoption is heavily influenced by infrastructure, policies, and, crucially, workforce skills and education [9]. While AI-enabled methods are becoming increasingly important for automation and decision support [10], [11], a common implication is that Construction 4.0 requires learning environments that can accelerate the development of procedural and spatial skills under safe, repeatable conditions, especially for complex construction processes.

B. VR in AEC: from collaborative systems to training and execution support

VR applications in AEC include both (i) collaborative design and coordination and (ii) training/education for construction tasks. Collaborative VR platforms based on OpenBIM have been extensively reviewed, highlighting interoperability and usability challenges and the need for more mature standards and workflows [12]. VR is widely recognized as a training tool in construction engineering education because it can simulate construction scenarios and facilitate experiential learning in complex procedural tasks. [2]. For construction management and site-related applications, scoping reviews also show broad uses of Augmented Reality (AR) and VR (e.g., safety, progress monitoring, quality/defect management, and training), reaffirming their importance as both a project-management and educational resource [13].

C. VR in education: effectiveness depends on instructional design and cognitive conditions

Across higher education, systematic reviews and meta-analyses report generally positive effects of immersive VR, but outcomes vary depending on implementation quality, learning design, and task-technology fit [3], [5]. Experimental evidence also warns that immersive VR can increase presence while decreasing learning if cognitive load is not managed [6]. Recent theory, e.g., the Cognitive Affective Model of Immersive Learning (CAMIL), views VR learning as a cognitive-affective process influenced by immersion and interactivity, in which presence, agency, and instructional guidance can either facilitate or hinder learning outcomes based on design choices [7]. Therefore, assessments of VR training should include both perceived learning conditions and objective performance data.

D. VR in construction management education and BIM-VR integration

In construction management education, integrating BIM with VR is proposed as a structured curriculum strategy to enhance students' understanding of construction concepts through immersive 3D learning experiences. Implementations highlight the importance of aligning with program outcomes, faculty training, and suitable infrastructure to ensure consistent educational results [14].

E. Gap and positioning of this study

Despite the increasing body of VR research in AEC, there is still limited empirical evidence concerning graduate learners performing prefabricated housing assembly tasks, particularly when using a mixed evaluation that combines (i) structured perceptions of experience quality and learning/transfer, and (ii) software-recorded performance metrics such as time, attempts, accuracy, and efficiency. This study addresses that gap by assessing an immersive, step-by-step VR workflow that simulates the structural assembly of a prefabricated house, triangulating perceived outcomes with objective performance indicators to develop actionable design insights for Construction 4.0-oriented education.

III. Immersive scenario design and VR montage sequence

Below, the VR experience implemented and the assembly task performed by the students are described. In particular, the educational purpose of simulation, the characteristics of the environment and the components involved, the sequence of assembly stages, and the main interactions required to complete the activity are presented to contextualize the subsequent analysis of perceptions and performance.

A. Educational purpose of the experience

The experience was designed as an immersive, guided training focused on developing two types of learning. On the one hand, procedural learning is associated with understanding the correct construction sequence for assembling prefabricated concrete elements that make up a house. On the other hand, spatial learning involves understanding the geometric relationships required to position and assemble components during installation correctly.

From the perspective of immersive learning design, it was prioritized that the student 'acts' within the environment, favoring presence and agency as relevant psychological conditions for learning in VR, in line with the literature on realistic behavior in immersive environments [1] and with recent models of immersive learning [7].

B. Virtual environment and assembly components

The immersive scenario represents the front of an outdoor construction site, with surrounding topography, an access road, and contextual elements typical of construction, such as signage, machinery, and support vehicles, to increase the plausibility of the context and promote behaviors closer to realistic execution during the task [1], [2]. In line with recommendations for educational VR, the environment prioritizes a functional visual composition (spatial clarity and references) over photorealism, to maintain focus on the assembly sequence and component identification [3].

In terms of resources, the environment includes storage areas for main components, such as foundation elements, wall panels, slabs, and beams, as well as a container dedicated to smaller connection components. These elements are presented in scale and arranged as the construction flow progresses.

To support spatial learning and reduce ambiguities in assembly, the system integrates consistent visual guidance signals: (i) floating labels that identify the active stage (e.g., “Foundations,” “Slab”), (ii) translucent geometries representing the target position of panels and beams before installation, and (iii) coupling markers and highlighted critical fixation points to guide the placement of anchors at element junctions. This set of visual aids aims to balance immersion and interactivity while managing cognitive load, in a manner consistent with immersive learning models that emphasize the role of instructional guidance and perceived control to achieve consistent educational outcomes [7].

Fig. 1 VR assembly environment - foundations stage.

C. Step-by-Step Assembly Workflow

The experience structures the process into a step-by-step, guided workflow, where each stage allows the next only once the planned assembly is finished. The stages include:

1) Foundations: Arrangement and alignment of the foundation elements to establish the system’s geometric reference.

2) Walls: Installation of wall panels on foundations, guided by translucent goal geometries to ensure position and orientation.

3) Wall-to-wall anchors: Placement of anchors at critical meeting points between walls, reinforcing the continuity and stability of the assembly.

4) Slab: Installation of the slab, consolidating the foundation, support, and geometric compatibility with the walls.

5) Wall-to-beam anchors: Incorporation of wall-beam connection anchors as a transition to the upper system, reinforcing the logic of structural continuity.

6) Beams: Installation and alignment of beams in their target position, verifying continuity and spatial coherence with the walls and supports.

7) Slab on beams: Finally, the upper slab is added on the beams, positioning it according to spatial guidance to ensure proper support, alignment, and system closure. This stage emphasizes understanding the entire system as an integrated whole, in which the slab acts as a rigid element and closure of the upper plane.

Throughout the process, the task is carried out using simple, repeatable actions (e.g., selecting, positioning, adjusting, and confirming), supported by guiding signals (stage labels, visual location references, and coupling markers). This step-by-step integration allows the participant to complete the structural assembly of the house as part of the VR assembly training task, reducing ambiguities in the assembly and promoting perceived control during execution [1], [7].

IV. Methodology

A. Participants and context

As mentioned, the study was conducted with final-semester graduate students enrolled in the master’s program in construction management and sustainability of the USS. The activity took place in the Industry 4.0 Laboratory in Concepción as part of the Construction 4.0 learning context. Participation was voluntary and anonymous, and no personally identifiable information was collected.

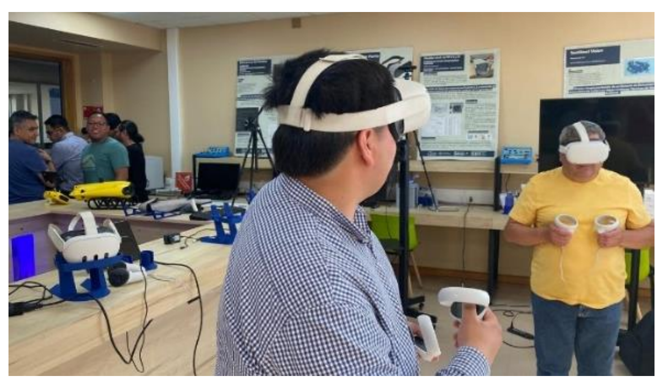

Fig. 2 Participants performing the VR assembly training task.

A total of n = 12 participants completed the VR task and generated software performance logs. The participant group should be understood as a criterion-based educational sample drawn from the specific graduate context in which the intervention was implemented. Accordingly, the sample was not intended to be statistically representative of broader student populations, but to provide contextually relevant evidence from learners directly aligned with the target training scenario.

The questionnaire included a pre-exposure screening to assess susceptibility to motion sickness and to identify safe participation options.

Ethical safeguards included informed consent, the right to withdraw at any time without academic penalty, and a safety protocol to address discomfort (such as dizziness, nausea, disorientation), in line with common considerations in immersive VR use for education and training.

B. Study design

The study was conducted as a single-session pilot test as part of a graduate learning activity. All participants followed a standardized procedure to ensure consistent exposure, instructions, and task conditions. The process was as follows:

1) Induction through an information session: Participants received a brief overview explaining the purpose of the activity and the step-by-step assembly task they needed to complete in VR.

2) Pre-exposure block: Participants completed an initial survey that included informed consent and a safe-use assessment focused on susceptibility to motion sickness and managing discomfort. They were told they could pause or stop the activity at any time without penalty and could choose alternative participation options if needed.

3) VR introduction and familiarization: Before starting the task, participants received a brief orientation on the VR setup, including basic interaction actions like selecting and placing items, to help reduce usability barriers unrelated to learning outcomes.

4) VR assembly training task Execution: Participants followed the guided assembly workflow outlined in Section III, progressing through each stage until the structure was fully assembled. Short breaks were permitted if participants experienced discomfort or fatigue.

5) Post-VR Experience Survey Block: Right after the VR session, participants completed the second part of the questionnaire, which included Likert scale items and open-ended questions. This step assessed the quality of the experience, perceived learning, and readiness for real-world scenarios.

6) Extraction and Linking of Objective Metrics: The VR software automatically recorded performance metrics during task execution. These records were exported and linked to each participant using an anonymous participant code, enabling triangulation between perceived results and objective performance.

This structured flow was designed to (i) ensure ethical and safe participation, (ii) reduce the confounding effects caused by initial unfamiliarity with the tool, and (iii) support an integrated analysis that combines self-reported measures with performance indicators recorded by the system.

C. Survey instrument

A structured questionnaire was used to collect (i) ethical consent and ensure safe participation before Head-Mounted Display (HMD) exposure, and (ii) to assess the perceived quality of the immersive learning experience immediately after the VR activity. The instrument was administered anonymously and designed to be completed in approximately 10-15 minutes. Responses in the main sections were measured on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree).

1) Pre-exposure block: informed consent and safety screening: Before starting the VR session, participants completed a brief block that included: (i) informed consent for voluntary and anonymous participation, (ii) self-assessment of susceptibility to motion sickness Q0-Q2, and (iii) a decision pathway for safe participation, such as an alternative non-HMD option or gradual exposure, along with a clear stop rule in case of symptoms like dizziness, nausea, disorientation, or visual discomfort.

This pre-exposure screening follows established guidelines in immersive VR research to proactively monitor and reduce motion sickness that can occur during these activities, which are usually assessed through standardized symptom questionnaires (e.g., SSQ) and structured exposure protocols [15].

2) Post-experience block: quality of the perceived experience and value of the learning: Immediately after the VR task, participants answered five subscales and a few open-ended questions, following the structure outlined below:

(i) Usability and control: This part of the questionnaire includes 5 items, from question Q3 to question Q7, and covers the following topics: perceived ease of use, integration of key functions, predictability, learnability, and confidence when interacting with the system. Although the questionnaire is specific to this study, it focuses on the same core usability aspects as widely used usability tools such as the System Usability Scale (SUS), thereby supporting comparability with previous Human-Computer Interaction (HCI) research [16].

(ii) Immersion and learning conditions: This part includes 9 items, from question Q8 to question Q16, and covers the following topics: presence (“being there”), intuitive and familiar interaction, controller precision, clarity of instructions, perceived visual order, enjoyment and motivation, HMD comfort, and perceived physical safety during movement.

(iii) Perceived learning: This section contains 7 items, from question Q17 to question Q23, including perceived understanding of the procedural sequence, spatial comprehension (positions/interferences/restrictions), ability to detect errors and critical points, understanding of rationale (“why”), perceived complementarity with prior course content, ability to explain the procedure using appropriate terminology, and perceived retention.

(iv) Transfer, self-efficacy, and intention to use: 6 items (Q24-Q29) evaluate the perceived willingness to apply learning in course tasks, confidence in performing similar tasks in real or semi-real contexts, confidence in adapting to changing scenarios, willingness to reuse VR, recommendation to continue the activity, and perceived value compared to traditional methods.

(v) Overall evaluation and qualitative feedback: To complement the multidimensional Likert subscales, the questionnaire included a final overall single-item assessment (Q30) of the VR experience on a scale from 1 to 7, summarizing their overall perception of the activity as a learning tool; and a brief qualitative component that gathered improvement needs and perceived educational value.

D. Performance metrics obtained from the software

During task execution, the VR application automatically exported participant-level performance logs, enabling objective measurement of procedural performance using time- and error-based indicators commonly used to assess VR training effectiveness and skill transfer [17], [18].

The exported metrics were:

(i) Completion time: the total number of minutes needed to finish the immersive assembly task.

(ii) Total attempts: the total number of system-recorded action attempts during the task.

(iii) Successes: the number of attempts the system has classified as correct.

(iv) Errors: the number of attempts labeled as incorrect by the system.

(v) Accuracy rate: the proportion of correct attempts, calculated as shown in equation (1).

$$Accuracy = \frac{Successes}{Total\ attempts} \quad (1)$$

(vi) Response speed: attempt rate, computed as shown in equation (2).

$$Speed = \frac{Total\ attempts}{Min\_Time} \quad (2)$$

To enhance interpretability and facilitate comparisons, two additional derived indicators were also calculated:

(i) Efficiency (successes per minute). As shown in the following equation (3).

$$Efficiency = \frac{Successes}{Min\_Time} \quad (3)$$

(ii) Errors per minute: See equation (4).

$$Errors/min = \frac{failures}{Min\_Time} \quad (4)$$

Together, these measures include temporal performance, error-driven rework, and overall execution efficiency, aligning with established evidence that time and error metrics are sensitive indicators of VR training improvements and their transfer to real-task performance.

E. Data analysis plan

Survey responses and software log data were analyzed using a mixed-methods triangulation approach that combined descriptive statistics, exploratory associations, and qualitative thematic analysis.

1) Quantitative survey analysis: First, item-level responses were summarized using the mean, standard deviation, median, and minimum-maximum. Afterward, composite scores were calculated as the average of items within each questionnaire block. Internal consistency for each multi-item block was assessed using Cronbach's alpha (α), an exploratory indicator of internal consistency given the pilot nature of the sample size ($n = 12$), following recommendations for preliminary scale assessment in small or exploratory samples [16], [19].

2) Quantitative analysis of software reports: Software logs were summarized using the same descriptive statistics for time, attempts, successes, errors, accuracy, and response speed, and were supplemented with derived indicators: efficiency (successes/min) and errors per minute to enhance interpretability and comparability across participants.

To examine the relationship between perceived experience/learning and objective task performance, Spearman rank correlations were calculated between: (i) survey block scores and (ii) key performance metrics (time, attempts, accuracy, efficiency, errors/min). Results were considered exploratory effect-size evidence due to the study's pilot nature.

3) Qualitative analysis using open-ended questions: Responses to open-ended questions (Q31-Q32) were analyzed using inductive thematic analysis, which involved familiarization, initial coding, theme grouping, and refinement into a concise set of recurring themes. The results were presented as (i) the main themes with brief representative excerpts and (ii) a summary of theme frequency to highlight the most consistent improvement priorities and perceived learning contributions.

4) Integration: To combine results from different data sources, we will use a triangulation strategy that links quantitative data from surveys and software with qualitative findings. The integration will be carried out in three clear steps.

(i) Profile construction: Participants will be grouped into interpretable performance profiles (e.g., efficient, average, trial-and-error) using efficiency (successes per minute), accuracy, errors per minute, and completion time. This step provides a clear baseline of how well the task was performed.

(ii) Convergence and divergence mapping: For each profile, we will compare the average scores of the survey blocks. This procedure helps identify convergent patterns (e.g., high usability and high efficiency) and divergent patterns (e.g., high enjoyment but low accuracy), which is essential in immersive learning research where experience quality does not always translate into performance.

(iii) Explanatory linking through qualitative themes: Open-ended responses will be used to explain why convergence or divergence occurs. For example, themes such as interaction precision, clarity of visual cues, or physical comfort can be linked to specific performance outcomes, such as high errors per minute associated with "control sensitivity" complaints. The final interpretation will report these connections as evidence-based design implications.

This integration is presented through a compact joint display that shows: performance profile, survey block means, and dominant qualitative themes, creating a coherent narrative that links "what participants felt", "what they achieved", and "what should be improved".

V. RESULTS

This section presents the results following the analysis plan, encompassing descriptions, performance records, associations, qualitative themes, and integrated interpretation.

A. *Quantitative survey results*

All participants completed the post-experience Likert questionnaire. Based on this, the responses from the post-experience questionnaire (Likert scale 1-7) were consolidated into block scores, calculated as the average of their corresponding items: Usability a and Control (Q3-Q7), Immersion and Learning Conditions (Q8-Q16), Perceived Learning (Q17-Q23), and Transfer/Self-efficacy/Intention (Q24-Q29), along with the overall experience rating (Q30). Table 1 summarizes the descriptive statistics (mean, standard deviation, median, and range) and the internal consistency (Cronbach's alpha) for each block for the total sample (n = 12).

TABLE I
DESCRIPTIVE RESULTS BY QUESTIONNAIRE BLOCK (N = 12)

Block	k	Mean	SD	Median	Min-Max	α
Usability and Control	5	6.617	0.463	6.700	5.40-7.00	0.576
Immersion and Learning Conditions	9	6.249	0.845	6.667	4.00-7.00	0.874
Perceived Learning	7	6.175	0.866	6.500	4.43-7.00	0.893
Transfer/Self-efficacy/Intention	6	6.333	1.064	6.583	3.17-7.00	0.938
Overall rating	1	6.250	0.622	6.000	5.00-7.00	-

Internal consistency was high for Immersion, Perceived Learning, and Transfer, while Usability showed lower consistency (α = 0.576), a pattern typical of small-sample pilots and possibly reflecting diverse perceptions across usability aspects.

Regarding the item-level highlights, based on their means, the highest-rated items were rapid familiarization and confidence (Q6 = 6.92; Q7 = 6.83) and ease of use (Q3 = 6.75). The lowest-rated items were the comfort of the HMD (Q15 = 5.50) and the perceived ability to identify errors or critical points compared to traditional methods (Q19 = 5.50).

Regarding the global rating (Q30), responses concentrated between 6 and 7 (mean = 6.25), with only one participant rating the experience as 5.

B. *Software-recorded performance metrics*

Alongside the self-reported measures, objective logs automatically generated by the software during the VR assembly training task were examined. These logs record procedural performance through indicators such as completion time, correct answers, errors, total attempts, success rate, and response speed, as well as derived metrics of interest for comparing participants: efficiency (correct answers per minute) and errors per minute. Table 2 displays the descriptive statistics for these variables for the entire sample (n = 12).

TABLE II
DESCRIPTIVE RESULTS FROM SOFTWARE LOGS (N = 12)

Metric	Mean	SD	Median	Min-Max
Completion time	10.692	2.650	10.500	6.80-15.00
Successes	22.000	2.089	22.000	20-24
Errors	16.833	11.800	13.500	3-40
Total attempts	38.833	10.495	36.000	27-60
Accuracy	0.609	0.184	0.619	0.33-0.89
Response speed	3.652	0.463	3.635	2.88-4.42
Efficiency	2.198	0.666	2.013	1.33-3.53
Errors/min	1.453	0.785	1.348	0.44-2.86

Objective records indicated an average completion time of 10.69 minutes, with a range of 6.80 to 15.00 minutes. The average accuracy was 0.609, ranging from 0.333 to 0.889. The number of correct answers remained relatively stable, between 20 and 24, while the number of errors varied widely, between 3 and 40, suggesting that most participants could complete the task but differed in the number of failed attempts.

C. *Exploratory associations between perceptions and performance*

To examine how perceived experience quality and learning relate to objective task performance, we conducted an exploratory association analysis combining (i) correlation tables and (ii) simple visual inspections through scatterplots. Given the pilot sample size (n = 12), results are reported as effect-size evidence (Spearman's ρ), with p-values used only as supplementary information.

1) Associations within objective performance metrics: First, we examined associations among software-recorded indicators to understand the internal structure of performance, i.e., whether a longer time indicates careful execution or error-driven rework. Table 3 summarizes the strongest correlations, showing a consistent pattern in which higher accuracy and efficiency correlate with fewer errors per minute and fewer total attempts, while longer completion time mainly reflects increased rework rather than deliberate accuracy.

TABLE III
KEY SPEARMAN CORRELATIONS AMONG OBJECTIVE PERFORMANCE INDICATORS (N = 12)

Metric A	Metric B	Spearman ρ	p-value
Accuracy	Errors/min	-0,984	<0.001
Attempts	Accuracy	-0,949	<0.001
Accuracy	Efficiency	0,929	<0.001
Min_Time	Efficiency	-0,925	<0.001
Min_Time	Attempts	0,879	<0.001
Min_Time	Accuracy	-0,801	0.002

2) Associations between survey blocks and objective performance: Next, we assessed relationships between block-level survey scores and key objective indicators. The full correlation matrix is best presented as a heat map, as shown in Figure 3, to emphasize the magnitude of the effects.

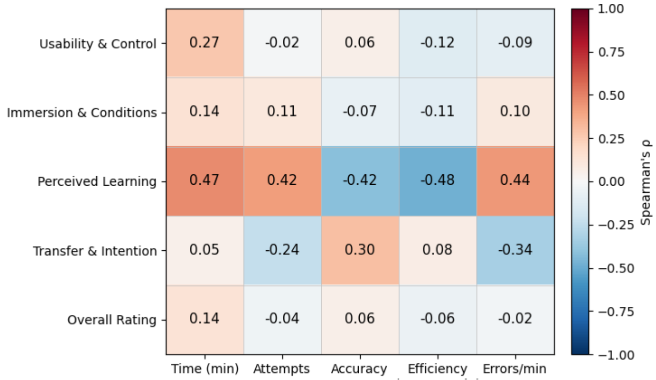

Fig. 3 Spearman Correlation Heatmap: Survey Blocks vs Performance Metrics.

The heat map indicates that, overall, the links between perception and objective performance are weak, though there are some specific trends. The strongest connection is in Perceived Learning, which is positively related to Time ($\rho = 0.47$), Attempts ($\rho = 0.42$), and Errors/min ($\rho = 0.44$), while it is negatively related to Efficiency ($\rho = -0.48$) and Accuracy ($\rho = -0.42$). This pattern suggests that those facing a more challenging task (more time, more attempts, and a higher error rate) reported feeling a greater sense of learning. In contrast, Usability and Control, Immersion and Conditions, Transfer and Intent, and Overall Rating show weak correlations with the objective indicators, suggesting that a high experience rating does not necessarily translate into better performance in this pilot.

3) Synthesis of association evidence: Across analyses, objective metrics showed strong and coherent internal coupling (accuracy-efficiency-errors/min), indicating that the software logs sensitively capture procedural performance. In contrast, perception-performance relationships were weaker and should be interpreted cautiously, given the pilot sample. Rather than treating non-significant correlations as null effects, we interpret these patterns as indicative of two coexisting phenomena common in immersive learning: (i) high satisfaction/immersion does not necessarily imply immediate proficiency, and (ii) perceived learning gains may reflect both mastery and the experience of challenge during task execution. These findings motivate the triangulated interpretation presented next and provide actionable targets for design improvement, e.g., interaction tuning and guided practice.

D. *Qualitative survey results*

Open-ended responses were obtained for both qualitative items (Q31-Q32). Thematic coding yielded a small set of recurrent themes that complement the quantitative findings by identifying (i) the most urgent design improvements and (ii) the learning value perceived by participants.

1) Improvement priorities (Q31): As summarized in Figure 4, the most frequently reported priority was improving physical space, mobility, and safety conditions, reflecting concerns about moving comfortably while wearing the headset. Participants also highlighted the need to increase task complexity and add activities, as well as to provide more practice or repetition, suggesting that the experience is valuable but could be enhanced through repetition and expanded scenarios. Additional themes mentioned include interaction precision (e.g., sensitivity when placing walls), immersion conditions (noise isolation), and visual aspects quality.

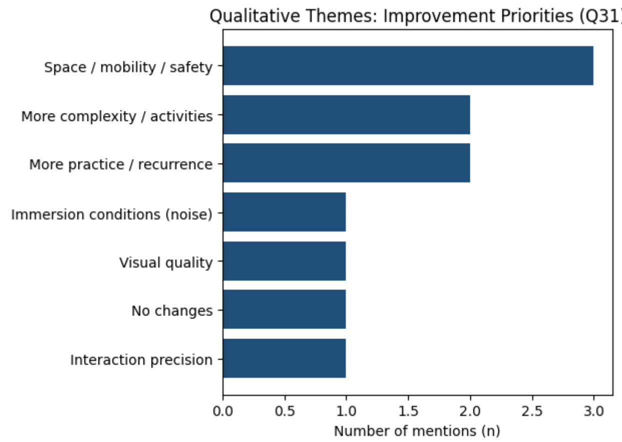

Fig. 4 Qualitative Themes: Improvement Priorities (Q31).

2) Most valuable contribution (Q32): As shown in Figure 5, participants most commonly highlighted improved understanding of the construction process/sequence and the value of VR/technology for learning. A smaller set emphasized the experience as a way to get closer to real work processes. Only one comment referred explicitly to teamwork, and one response expressed a critical view ("new tools"), indicating a minority divergence in perceived utility.

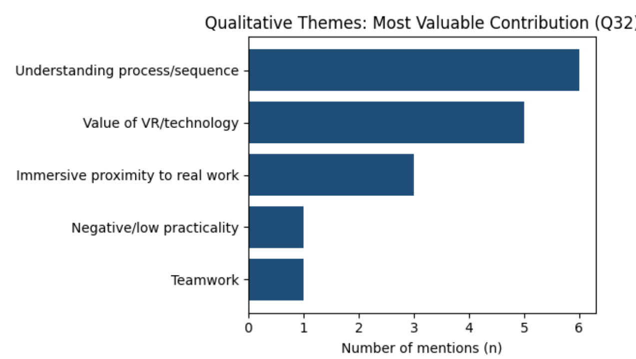

Fig. 5 Qualitative Themes: Most Valuable Contribution (Q32).

E. *Integrated findings*

To integrate evidence from multiple instruments, results were triangulated by aligning objective execution profiles (software logs) with (i) survey block scores and (ii) open-ended themes. Three performance profiles were identified based on efficiency and error-driven rework patterns: Efficient (high efficiency/accuracy, low errors per minute; n=4), Intermediate (moderate efficiency/accuracy; n=6), and Trial-and-error (low efficiency/accuracy, high errors per minute; n=2). The integration is summarized in Table 4 and can be further visualized in a radar/spider plot (Figure 6) to compare multidimensional perception patterns across profiles.

TABLE IV
SURVEY BLOCK MEANS BY PERFORMANCE PROFILE (LIKERT 1-7; N=12)

Profile	n	Usability and Control	Immersion and Conditions	Perceived Learning	Transfer and Intention	Overall Rating
Efficient	4	6.81	6.44	5.88	6.63	6.50
Intermediate	6	6.50	5.94	6.14	6.08	6.00
Trial-and-error	2	7.00	6.74	6.86	6.50	6.50

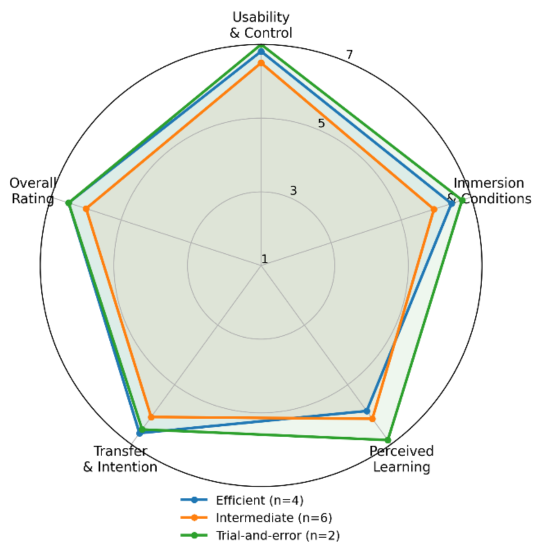

Fig. 6 Radar Plot: Survey Block Means by Performance Profile.

To complement the radar chart, Figure 7 highlights a clustered divergence between perception and performance across profiles, rather than a stronger overall correlation.

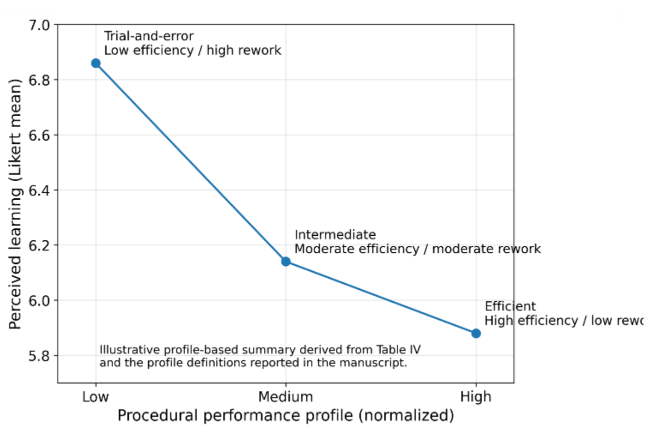

Fig. 7 Illustrative perception-performance divergence across execution profiles.

From the triangulation, we obtained the following key ideas:

(i) High perceived quality coexists with varied performance. All groups rated the experience positively (Overall ≈ 6.0-6.5), including the Trial-and-error group, showing strong acceptance even when actual performance was low. This finding highlights the importance of reporting objective logs alongside self-reports.

(ii) Perceived learning may reflect "productive struggle." The Trial-and-error profile reported the highest Perceived Learning (6.86) despite the lowest accuracy/efficiency, while the Efficient profile reported lower Perceived Learning (5.88). This result aligns with the correlation pattern observed in the heatmap (Section C2) and suggests that greater challenge and rework can boost perceived learning gains in short, single-session pilots.

(iii) Qualitative themes identify performance bottlenecks. The most common improvement requests, more physical space, mobility, safety, higher task complexity, more practice, and interaction precision, align with the wide variation in errors and attempts. For instance, "interaction sensitivity" and room-scale constraints offer plausible explanations for high errors per minute in the Trial-and-error profile, while "more practice" suggests that repetition could help stabilize performance across participants.

VI. DISCUSSION

This pilot study examined an immersive VR training experience for the structural assembly of prefabricated houses in a graduate Construction 4.0 course, combining self-reported perceptions with performance data recorded by software. Overall, participants reported high ratings in usability, learning conditions, perceived learning, and transferability, as shown in Table I, and the overall didactic rating was also high. These results align with broader evidence that VR can be perceived as an effective and engaging training modality when the match between the task and technology is appropriate [2], [3].

A. Interpretation of perceived experience quality (RQ1-RQ2)

The high average scores across blocks indicate that users found the experience easy to use, motivating, and supportive of procedural understanding and transfer readiness. This finding aligns with the overall trend in higher-education VR reviews and meta-analyses, which report positive effects but highlight that outcomes heavily depend on design choices and instructional alignment [3], [5]. It is worth noting, as shown in Table I, that the usability block had a lower internal consistency (α) in this pilot study with a small sample. This result may be due to varied usability aspects, for example, controller precision versus learnability, suggesting that future versions should refine usability items or compare them against standardized tools like the SUS [16].

B. What the objective logs reveal about procedural performance

Objective performance logs indicated that most participants could complete the workflow, but with considerable variability in errors and attempts (Table II). The strong correlation among accuracy, efficiency, attempts, and errors/min (Table III) suggests that longer completion times were mainly due to error-driven rework rather than slower, more careful execution. This finding underscores the importance of combining time and error metrics, which have consistently proven to be sensitive indicators in VR skills training and performance evaluation [17], [18].

C. Perception-Performance Gap (RQ3)

A key contribution of this study is the triangulated finding that high satisfaction and immersion do not necessarily indicate high proficiency. Correlations between survey results and objective indicators were generally weak, with the clearest trend showing that Perceived Learning increased with time, attempts, and errors, and decreased with efficiency and accuracy (Figure 3). This limited convergence is theoretically plausible because self-reported measures primarily capture cognitive-affective responses to the experience, whereas software logs reflect early procedural efficiency. In a brief initial exposure, these dimensions should not be expected to align fully. This pattern is consistent with prior evidence showing that immersive VR can intensify presence without necessarily improving immediate learning when cognitive demand is not adequately regulated [6]. Beyond confirming CAMIL, the present findings nuance it in the context of short, single-session VR training: cognitive-affective gains such as engagement, agency, and perceived learning may emerge earlier than measurable procedural efficiency. This result helps explain the temporary perception-performance divergence observed here, in which participants may feel they learned more precisely because the task required repeated adjustments and error correction. From an educational standpoint, this pattern is compatible with productive struggle, in which initial difficulty can support deeper procedural understanding before execution becomes stable and efficient.

D. Design and pedagogical implications for Construction 4.0 training

The qualitative themes provide actionable guidance for improving learning outcomes, safety, and comfort: participants emphasized more physical space, mobility, safety measures, additional practice and repetition, increased task complexity, and more precise interaction.

From an instructional design perspective, the results suggest three immediate priorities: (i) a brief pre-task familiarization phase to reduce interface-related noise; (ii) structured repetition across trials to transform early productive struggle into stable performance; and (iii) refinement of interaction precision and visual scaffolds to reduce avoidable error-driven rework. In practical terms, this implies optimizing room-scale setup and noise conditions, adjusting snapping/tolerance thresholds for assembly tasks, and using clearer stage cues, staged feedback, and adaptive hints. These refinements align with research on immersive learning that emphasizes guided design over immersion alone [3], [7]. In Construction 4.0 contexts, they support the development of safe, standardized, and evidence-based training environments for procedural and spatial skill acquisition [9].

E. Limitations and future work

This study should be interpreted as an exploratory, single-session pilot conducted in an authentic graduate learning setting. The reduced sample size (n = 12) limits statistical generalization and restricts the stability of subgroup and correlation patterns. Therefore, the results are better understood as preliminary evidence of possible relationships between perceived experience and procedural performance, rather than as confirmatory findings.

Future research should test whether the perception-performance divergence observed here persists or narrows under comparison conditions and across repeated exposure. This finding calls for designs that include comparison groups (e.g., video-based instruction, non-immersive 3D, or traditional guided practice), pre- and post-learning measures, and longitudinal follow-up capable of distinguishing novelty effects from actual skill consolidation. Finally, future work should improve system instrumentation and design by logging stage-level performance metrics (time and errors by phase), incorporating adaptive guidance (e.g., contextual hints and progressive scaffolding), and systematically tracking comfort (e.g., SSQ-based symptom monitoring). Collectively, these steps would support the iterative development of more effective, safe, and scalable immersive learning experiences for construction processes.

From a scalability perspective, the proposed VR workflow appears technically replicable because the environment, task sequence, and logging system can be reproduced under standardized conditions. However, pedagogical scalability remains unconfirmed, since this pilot does not establish whether similar perception-performance patterns would remain stable across larger cohorts, repeated implementations, or different educational contexts.

Although this pilot focused on prefabricated housing assembly, the didactic logic and triangulated evaluation framework may be extended to other immersive construction tasks involving procedural sequencing, spatial reasoning, model-based decision-making, and multidimensional user-experience demands. Relevant examples include MEP coordination, technical model review, construction control and operation, and documentation and information retrieval support, all of which provide suitable contexts for extending the present framework across construction programs and institutions [20].

VII. Conclusions

This paper presented an exploratory pilot evaluation of an immersive VR workflow for teaching the structural assembly of a prefabricated house in graduate education. It used a mixed-methods strategy, combining learner perceptions (experience quality, perceived learning, and transfer intention) with software-recorded performance logs (time, attempts, errors, accuracy, and efficiency). The findings directly address the three research questions and clarify how immersive learning benefits and limitations manifest in construction-process training.

Regarding the research question RQ1, graduate students consistently reported high perceived experience quality, including usability/control and immersion/learning conditions, with a strong overall didactic rating. For RQ2, participants also reported high perceived learning and transfer readiness, and qualitative feedback confirmed that the most significant value was improved understanding of the construction process and sequence. These findings support that immersive VR can be effectively integrated into construction education when supported by structured instructional design and measurable performance metrics [2], [3].

Crucially, RQ3 revealed the paper's most notable insight: positive experience ratings did not necessarily lead to higher objective proficiency. Performance logs showed significant variation in error-driven rework, with strong internal connections among accuracy, efficiency, attempts, and errors/min, indicating that longer completion times mainly reflected rework rather than careful execution. Meanwhile, the perception-performance heatmap indicated that Perceived Learning tended to increase with time, attempts, and errors, and decrease with efficiency and accuracy, suggesting a "productive struggle" pattern in a single-session exposure. This nuance aligns with evidence that immersive VR can boost presence without guaranteeing learning gains if cognitive demand and interaction constraints are not properly managed [6], and it supports the CAMIL perspective that learning outcomes depend on how immersion and interactivity influence cognitive-affective processes under instructional guidance [7].

Beyond reporting satisfaction, this study demonstrates a practical, replicable way to evaluate VR for construction-process education through triangulation: (i) validated block-based perception measures, (ii) objective system logs that accurately capture procedural performance, and (iii) qualitative themes that identify bottlenecks and suggest improvements. The work thus advances the field from "VR is promising" to "VR is measurable," providing actionable evidence on how to design, assess, and enhance immersive VR training for prefabricated housing assembly and, potentially, for other construction education tasks that involve procedural, spatial, and model-based decision-making.

REFERENCES

[1] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3549–3557, Dec. 2009, doi: 10.1098/rstb.2009.0138.

[2] P. Wang, P. Wu, J. Wang, H.-L. Chi, and X. Wang, "A Critical Review of the Use of Virtual Reality in Construction Engineering Education and Training," *Int. J. Environ. Res. Public Health*, vol. 15, no. 6, p. 1204, Jun. 2018, doi: 10.3390/ijerph15061204.

[3] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," *Comput. Educ.*, vol. 147, p. 103778, Apr. 2020, doi: 10.1016/j.compedu.2019.103778.

[4] M. C. Howard, M. B. Gutworth, and R. R. Jacobs, "A meta-analysis of virtual reality training programs," *Comput. Human Behav.*, vol. 121, p. 106808, Aug. 2021, doi: 10.1016/j.chb.2021.106808.

[5] M. Coban, Y. I. Bolat, and I. Goksu, "The potential of immersive virtual reality to enhance learning: A meta-analysis," *Educ. Res. Rev.*, vol. 36, p. 100452, Jun. 2022, doi: 10.1016/j.edurev.2022.100452.

[6] G. Makransky, T. S. Terkildsen, and R. E. Mayer, "Adding immersive virtual reality to a science lab simulation causes more presence but less learning," *Learn. Instr.*, vol. 60, pp. 225–236, Apr. 2019, doi: 10.1016/j.learninstruc.2017.12.007.

[7] G. Makransky and G. B. Petersen, "The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality," *Educ. Psychol. Rev.*, vol. 33, no. 3, pp. 937–958, Sep. 2021, doi: 10.1007/s10648-020-09586-2.

[8] G. B. Petersen, G. Petkakis, and G. Makransky, "A study of how immersion and interactivity drive VR learning," *Comput. Educ.*, vol. 179, p. 104429, Apr. 2022, doi: 10.1016/j.compedu.2021.104429.

[9] S. V. Jaiswal, D. V. L. Hunt, and R. J. Davies, "Construction 4.0: A Systematic Review of Its Application in Developing Countries," *Applied Sciences*, vol. 14, no. 14, p. 6197, Jul. 2024, doi: 10.3390/app14146197.

[10] S. K. Baduge *et al.*, "Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications," *Autom. Constr.*, vol. 141, p. 104440, Sep. 2022, doi: 10.1016/j.autcon.2022.104440.

[11] R. Valdebenito and E. Forcael, "Integrating Artificial Intelligence and BIM in Construction: Systematic Review and Quantitative Comparative Analysis," *Applied Sciences*, vol. 15, no. 23, p. 12470, Nov. 2025, doi: 10.3390/app152312470.

[12] D. Ververidis, S. Nikolopoulos, and I. Kompatsiaris, "A Review of Collaborative Virtual Reality Systems for the Architecture, Engineering, and Construction Industry," *Architecture*, vol. 2, no. 3, pp. 476–496, Jul. 2022, doi: 10.3390/architecture2030027.

[13] M. Albahbah, S. Kıvrak, and G. Arslan, "Application areas of augmented reality and virtual reality in construction project management: A scoping review," *Journal of Construction Engineering, Management & Innovation*, vol. 4, no. 3, Sep. 2021, doi: 10.31462/jcemi.2021.03151172.

[14] S. Y. Ghanem, "Implementing virtual reality - building information modeling in the construction management curriculum," *Journal of Information Technology in Construction*, vol. 27, pp. 48–69, Jan. 2022, doi: 10.36680/j.itcon.2022.003.

[15] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness," *Int. J. Aviat. Psychol.*, vol. 3, no. 3, pp. 203–220, Jul. 1993, doi: 10.1207/s15327108ijap0303_3.

[16] A. Bangor, P. T. Kortum, and J. T. Miller, "An Empirical Evaluation of the System Usability Scale," *Int. J. Hum. Comput. Interact.*, vol. 24, no. 6, pp. 574–594, Jul. 2008, doi: 10.1080/10447310802205776.

[17] N. E. Seymour *et al.*, "Virtual Reality Training Improves Operating Room Performance," *Ann. Surg.*, vol. 236, no. 4, pp. 458–464, Oct. 2002, doi: 10.1097/00000658-200210000-00008.

[18] T. P. Grantcharov, V. B. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, and P. Funch-Jensen, "Randomized clinical trial of virtual reality simulation for laparoscopic skills training," *Journal of British Surgery*, vol. 91, no. 2, pp. 146–150, Jan. 2004, doi: 10.1002/bjs.4407.

[19] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *Int. J. Med. Educ.*, vol. 2, pp. 53–55, Jun. 2011, doi: 10.5116/ijme.4dfb.8dfd.

[20] M. García, J. Requesens, and S. Cano, "User Experience Evaluation Methods in Mixed Reality Environments," 2024, pp. 179–193. doi: 10.1007/978-3-031-61281-7_12.