

Evaluating Aircraft Cabin Design in Virtual Reality: The Role of Environmental Context

Jessica Herzig¹, Eike Langbehn², Tessa T. Taefi², Justin Mittelstädt³, and Jörn Biedermann¹

¹German Aerospace Center (DLR e.v.), Institute of System Architectures in Aeronautics, Hein-Saß-Weg 22, 21129 Hamburg, Germany

²Hamburg University of Applied Sciences, Faculty of Computer Science and Digital Society, Berliner Tor 5, 20099 Hamburg, Germany

³German Aerospace Center (DLR e.V.), Institute of Aerospace Medicine, Sportallee 54A, 22335 Hamburg, Germany

ABSTRACT

As virtual reality (VR) becomes more integrated into collaborative design and development processes, it is essential to understand how visual context influences user experience and evaluation. Since visual perception is highly subjective and varies across stakeholder groups, the design of the visual environment must be carefully considered, as it directly shapes the sense of presence. However, the role of environmental context in VR-based design evaluations remains largely unexplored. This paper investigates the influence of virtual contexts on evaluations of aircraft cabins. The goal is to identify combinations of cabin visualisation and environmental design that promote positive and consistent user experiences. Our VR user study empirically compares multiple environmental contexts and aircraft visualisations and assesses perceived appropriateness, task effectiveness, and distraction. Specifically, the study examines hangar, airport, and open-sky contexts in combination with different levels of cabin visualisation. The results suggest that a closed cabin combined with a sky environment is the most conducive setting. Overall, the findings demonstrate that virtual context significantly impacts well-being, concentration, and evaluation consistency, highlighting its crucial role in VR-based cabin assessment.

Keywords: Virtual reality, Cabin design, Spatial design, User experience, Human factors

INTRODUCTION

Aircraft cabin development requires complex multi-criteria design decisions. Economic considerations related to efficient space utilisation must be balanced against passengers' expectations regarding comfort and aesthetics, the functional requirements of cabin crew, and strict regulatory safety constraints (Moerland-Masic et al., 2022; Wise et al., 2010).

In this context, virtual reality (VR) serves as a technological medium for implementing cabin models and facilitating the design process. Within VR,

immersive virtual environments (IVEs)¹ enable stakeholders to experience and evaluate spatial configurations, usability and comfort in an immersive, contextualised setting. This reduces the need for physical prototypes in the early stages of design and allows end users to be involved from the outset. VR-based IVEs support more agile and informed decision-making in cabin development by enabling rapid iteration and contextual evaluation prior to physical implementation (Kozhevnikov and Garcia, 2011; Moerland-Masic et al., 2022). As part of the broader integration of digital tools in the aerospace industry, VR as an enabling technology has thus emerged as a key medium for accelerating the development and evaluation of innovative cabin design solutions (Herzig et al., 2025; Reimer et al., 2024).

Compared to other immersive technologies, such as on-screen visualisation or CAVE (Cave Automatic Virtual Environment) systems, HMD-based VR systems offer distinct advantages: they enable a higher level of immersion through stereoscopic depth perception and natural head-tracked interaction, require less physical infrastructure, and allow flexible deployment across distributed teams. These characteristics make HMD-based VR particularly suitable for iterative design evaluations and interdisciplinary decision-making in industrial contexts (Havig et al., 2011; Khairul et al., 2025; Kozhevnikov and Garcia, 2011).

However, the effectiveness of IVEs depends not only on underlying VR technology, but also on the quality of their spatial and experiential design. Although IVEs can convincingly recreate spatial and visual experiences, they may still restrict the user's creative freedom of thought and action compared to the real world in certain cases. In this context, the architecture and design of virtual spaces are of great importance, as the sense of presence in IVEs largely depends on perception and cognitive engagement being supported by the design of the spaces (Oxman, 2011).

While IVEs are widely used in collaborative design processes, the question of how to design virtual spaces to effectively support creative collaboration has remained open for a long time (Maher, 2011). Although recent studies demonstrate that environmental characteristics can influence presence, emotional state, and attention, actionable design guidance for creativity and collaboration within VR-based systems remains limited (Bartl et al., 2022).

This study investigates how the design of IVEs and aircraft visualisations impacts stakeholders' evaluation of cabin concepts. Rather than identifying a single optimal solution, the study takes an exploratory approach, collecting empirical evidence on how different environmental contexts and representation choices influence user evaluations. By systematically examining and structuring these effects, the study provides a data-driven foundation for the rationalisation of future work and subsequent collaborative design investigations using VR as an evaluation medium.

¹In this paper, IVE refers to a spatially designed, immersive virtual environment experienced through a head-mounted display (HMD)-based VR system.

RELATED WORK

Spatial Design

The physical environment plays a fundamental role in shaping human behaviour, structuring how actions are enabled, constrained and perceived within broader social and cultural contexts (Gibson, 1979; Waterlander et al., 2015). Spatial design addresses this relationship by deliberately configuring elements such as spatial layout, furnishings, lighting and atmosphere in order to influence how environments are experienced. Rather than directly prescribing behaviour, spatial design creates conditions that support activities such as concentration, learning, creativity and collaboration (Porqueddu, 2025).

Research on creative and learning environments suggests that the impact of spatial design stems from the interaction of spatial, functional and atmospheric factors operating at various levels. These factors range from architectural structure and spatial organisation to room layout, materials, and sensory qualities. This suggests that creative engagement is shaped by a space's overall experiential context rather than by individual design elements (Thoring et al., 2020).

As these principles have primarily been studied in physical environments, their relevance and transferability to virtual spaces is not well understood. Immersive technologies such as VR allow spatial and experiential conditions to be represented and manipulated in controlled settings. This provides new opportunities to investigate how environmental design influences perception and behaviour beyond physical constraints (Neo et al., 2021; Waterlander et al., 2015).

Immersive Virtual Environments

The human understanding of reality is largely shaped by visual perception, which accounts for around 80% of the sensory information we process (Man and Olchawa, 2018). Vision conveys more content and meaning than other senses and plays a key role in creating a sense of presence. This makes the deliberate design of IVEs in VR, particularly important (Chan, 2011; Kozhevnikov and Garcia, 2011).

IVEs are transforming design processes by replacing conventional representations, such as sketches, with dynamic, full-scale digital models. Although human perceptual mechanisms remain unchanged, the virtual context introduces new challenges and opportunities for understanding spatial relationships and environmental effects. In this context, the interplay between representation and perception is crucial, as the quality of design decisions in IVEs depends on designers' ability to translate ideas into external representations and interpret them effectively (Chan, 2011). Furthermore, emerging interaction technologies, such as tangible interfaces, are changing the way virtual spaces are perceived and interacted with in design contexts, thereby influencing designers' creative thinking and actions (Maher, 2011).

Therefore, the design of IVEs should meaningfully and holistically address “human perception, cognition and emotion” (Oxman, 2011:43). The goal of immersive and interactive technologies is to create a strong sense of presence while enabling smooth, intuitive interaction between participants. In this context, the architecture of these environments is of central importance, as it not only makes them responsive and flexible, but also supports their adaptation to different design styles and goals (Maher, 2011; Oxman, 2011).

Extended Reality Collaborative Design

In the course of technological development, new methods have emerged within IVEs that not only involve designers in virtual creative processes, but also integrate other stakeholders into the design process (Santhosh and De Crescenzo, 2023). The objective is to ensure that the newly designed outcomes align with user needs by placing their experiences and requirements at the centre of the development process, treating them as “experts on their own experiences” (Visser et al., 2005:10). This user-centred design approach of so-called co-creation incorporates “any act of collective creativity, i.e., creativity that is shared by two or more people” (Sanders and Stappers, 2008:6), while co-design refers to “the creativity of designers and people not trained in design working together in the design development process” (Sanders and Stappers, 2008:6).

Extended Reality (XR) — as an umbrella term encompassing Virtual, Augmented, and Mixed Reality — has significantly expanded the potential of collaborative IVEs (Rauschnabel et al., 2022; Skarbez et al., 2021). The use of these technologies as “generative tools” (Sanders and Stappers, 2019:78) not only enables diverse stakeholders to engage interactively in creative development processes independently of their location (El-Jarn and Southern, 2020), but also opens up new opportunities for interdisciplinary collaboration and learning (Schmohl, 2021). Similar to the work of Oxman, Maher and Chan, El-Jarn et al. argue that XR co-design processes suggest to “find a space” in which collaboration is strengthened in order to promote co-design. According to this perspective, the technology should help to overcome spatial distances and strengthen creative expression to improve the quality of design processes in the early stages (El-Jarn and Southern, 2020). According to Dorta et al., the aim is to design a suitable “representational ecosystem” that promotes a structured verbal exchange in order to improve the participation of non-designers in co-design processes (Dorta et al., 2019).

XR Co-Design Environments for Aircraft Cabins

Despite the growing exploration of XR-based co-design approaches across different domains, few studies have examined how XR co-design environments should be designed for the collaborative development of aircraft cabins. The following examples demonstrate how design choices in XR environments can influence spatial perception, sense of presence, and the practicality of collaborative design outcomes.

In their XR co-design study on developing a seat layout for a hydrogen aircraft cabin, Herzig et al. (2025) demonstrated that, despite achieving

high levels of immersion, the sense of actual presence was relatively weak. Participants, including designers, engineers, and non-experts (representing passengers), expressed a desire for a more detailed cabin representation, for example integrating emergency exits, cabin systems, and other central elements, to enable a realistic seating concept creation. This suggests that while a simply designed environment can establish spatial perception, it is insufficient to evoke a convincing sense of presence in the virtual space.

Another example of the importance of targeted design in XR co-design environments in the context of cabin environments can be seen in the study by Cornelje et al. (2024). In a workshop, emergency physicians collaboratively designed the interior of a virtual helicopter model. The model was half open and presented in low-poly style and could be edited using the 3D sketch tool 'Gravity Sketch'. As a conceptual design result, a two-storey lounge cabin was developed. However, after further review and the identification of additional structural issues, the project lead concluded that the original design was unfeasible and had to be scaled down to a single level. The half-open representation of the cabin may have distorted the participants' sense of space, while the low-poly visualisation made the overall impression appear more conceptual than realistic. Together, these factors could have prevented the participants from fully immersing themselves in the virtual reality and perceiving it as a work-related environment.

RESEARCH DESIGN

This study is part of the ongoing research at the German Aerospace Center (DLR e.V.) on advancing XR-based co-design methods for the collaborative development and evaluation of aircraft cabins. Within this broader context, the present study focuses specifically on the VR-based component, examining how the design of virtual spaces and aircraft representations influences cabin perception.

By concentrating on individual evaluations in a user study, the influence of group dynamics such as conformity or dominance is minimised, allowing subjective perceptions to be captured in a controlled manner (Diehl and Stroebe, 1987; Paulus and Brown, 2007). This approach provides initial reference values and supports the systematic collection of empirical data to identify influential as well as negligible or potentially limiting environmental and representational factors. Through inductive analysis, the study derives initial explanatory insights that inform subsequent investigations of XR-based co-design settings.

Based on this approach, the following research questions guide this investigation:

- RQ1:** How do different virtual settings compare in terms of perceived evaluation quality for evaluating aircraft cabin concepts in VR?
- RQ2:** Which virtual environmental characteristics are perceived as relevant, negligible, or limiting during the VR-based evaluation of aircraft cabin concepts?

To address these research questions, a VR user study was conducted to investigate the influence of virtual context design on the evaluation of a cabin layout. This evaluation was conceptualised as a latent construct and measured using multiple perceptual indicators to capture perceived task effectiveness, appropriateness and distracting elements.

The study was approved by the ethics committee of the DLR e.V.

Experimental Setup

The VR application was developed using Unreal Engine 5 and deployed on a Meta Quest 3 headset. It was executed on an Alienware m15 R6 laptop equipped with an Intel Core i7-11800H CPU, an NVIDIA GeForce RTX 3070 Laptop GPU, and 32 GB of RAM, running Windows 11. The application made it possible to teleport around a hydrogen aircraft with an innovative cabin concept in a realistic high dynamic range image (HDRI) environment.

While the environment and external aircraft representation varied, the underlying cabin concept remained constant in order to isolate the effects of the virtual context on the evaluation.

A combined locomotion system was employed to navigate within the virtual environment. Teleportation was utilised for localized locomotion using the right-hand controller, while real-time head tracking was used to translate real-world head movements into corresponding changes in the virtual environment. Additionally, the rotation around one's own axis in 45° increments (snap turn) was controlled using the left-hand joystick.

Although teleportation does not provide a natural sense of movement, it was chosen over continuous locomotion and the point-and-tug method due to its proven effectiveness in reducing motion sickness, and thereby minimizing potential biases in the collected results (Prithul et al., 2021).

To determine the influence of the environment and the representation of the aircraft model on the perception of the cabin concept, three context-related HDRI environment maps were used in the application. Based on spatial design research indicating the positive influence of openness and access to natural views on creative work environments, the concept of open views was implemented through sky and airport scenery, taking daylight conditions into account (Thoring et al., 2020). In addition to these outdoor contexts, a hangar was selected as a contrasting indoor environment to enable a systematic comparison of different spatial impressions. The aircraft model was represented using a closed cabin, a semi-transparent version and a half-open version. The closed version corresponds to the conventional standard and was therefore essential. The idea of a semi-transparent cabin and fuselage was inspired by Oxman's work on form and content, in which semi-transparent walls are said to strike "a balance between privacy and visual connection" with the surroundings (Oxman, 2011). Additionally, the objective was to render the cabin wall structurally and schematically recognizable without significantly impeding the perception of the outside world. The third form of visualisation, based on the work of Cornelje et al. (2024) was a longitudinal, half-open aircraft model. We chose this

third visualisation to investigate the influence of such a model on altered perceptions of space.

The environment and the aircraft model were combined within the application to create a 3×3 matrix of nine distinct settings. In the following text, the term ‘setting’ refers to the conceptual combination of environment and aircraft representation. To ensure clarity and reproducibility in the analysis, each setting is assigned a numerical identifier ranging from Level 1 to Level 9 (see Figure 1).

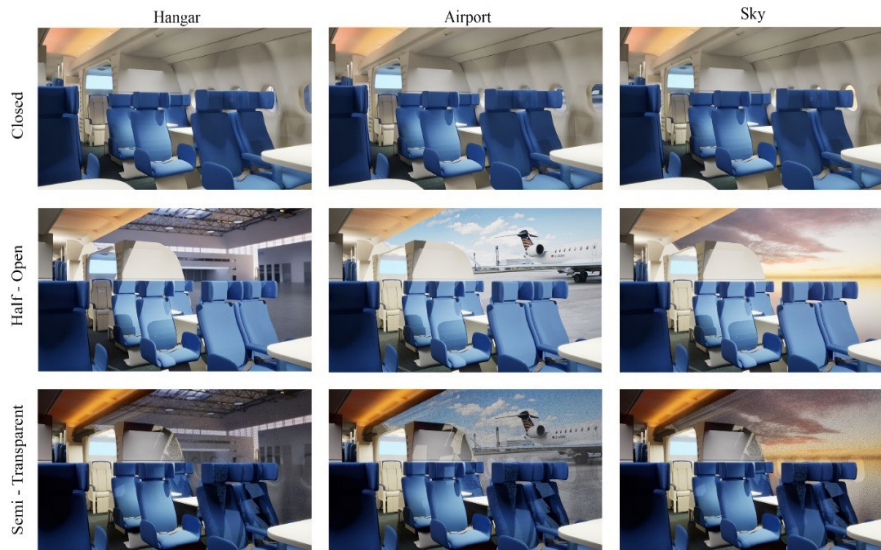


Figure 1: Screenshots of the VR application with nine settings (Level 1-9), which vary in terms of aircraft visualisation and virtual environment.

A random sequence of the nine combinations was generated for each participant using random allocation without replacement. This ensured that each combination appeared exactly once per participant, thereby minimizing sequence effects and potential biases. The order of presentation was documented in the application alongside the participants’ names.

The participants remained in the same location throughout all level transitions to preserve their movement patterns within the cabin.

Measurements

To improve usability and user experience (UX), and to better address user needs, it is essential to systematically measure and analyse in order to optimise in an informed and targeted way (Krug, 2013; Wang and Wang, 2011). A mixed-methods approach was employed to combine quantitative measures, which support comparability and statistical analysis, with qualitative responses that provide contextual insight into participants’ perceptions and interpretations. This approach addresses the complementary strengths and limitations of each method (Pilcher and Cortazzi, 2024).

The aim was to translate the subjective impressions, particularly with regard to UX, into measurable results, thereby gaining a comprehensive understanding of the usability and impact of IVEs. For the quantitative assessment, usability metrics were employed to capture key aspects of user-system interaction and support the systematic analysis of subjective experiences (Tullis and Albert, 2013).

In addition to the quantitative measures, qualitative data was collected via audio recordings of the VR sessions. All sessions were transcribed verbatim. This qualitative material was then used to provide contextual interpretation and illustrative support for the quantitative findings.

Representative participant statements were selected to exemplify recurring themes related to environmental perception, distraction and evaluation strategies. The qualitative component served an explanatory role, providing insight into how participants interpreted and articulated their experiences.

As impressions of the nine different settings can fade quickly, the survey was divided into two questionnaires. All responses were recorded using a 7-point Likert scale ranging from ‘Strongly Disagree’ to ‘Strongly Agree’, as this type of scale is considered to be particularly reliable and valid (Kusmaryono et al., 2022). The neutral response option ‘4’ was not labelled to encourage participants to engage more consciously with the statements and reduce their tendency to choose an indifferent middle position. The aim was to record clear agreement or disagreement, thus increasing the variance of the results (Froman, 2014).

To avoid disrupting participants’ immersion, the first questionnaire was administered verbally within the virtual environment immediately after each setting, and the responses were recorded by the experiment supervisor. Participants completed the second questionnaire themselves after the entire VR session.

Both questionnaires were written in English to ensure linguistic consistency and to include participants with different native languages.

VR-Session Questionnaire

The VR session questionnaire consisted of three concise statements, each of which was presented after a different experimental setting. Participants were asked to rate each statement to capture their immediate evaluation experience.

Q1 I was able to evaluate the cabin effectively in this setting.

Q2 The environment and aircraft visualisation felt appropriate for the cabin evaluation.

Q3 I noticed elements that made evaluation more difficult.

These statements assessed three complementary aspects of evaluation quality, which are based on well-established research into usability, contextual fit and cognitive load.

Q1 focused on effectiveness, reflecting the participant’s perceived ability to evaluate the cabin design and successfully complete the evaluation task,

in line with the established usability concept of task effectiveness (ISO 9241-11) (Nik Ahmad and Hasni, 2021).

Q2 addressed the appropriateness of the virtual environment and aircraft visualisation in relation to the evaluation task. This concept is closely related to “appropriate to the particular context” (ISO 9241-210) and ecological validity in human-centred design (Gibson, 1979; International Organization for Standardization., 2019).

Q3 targeted perceived distracting elements, reflecting extraneous perceptual or cognitive factors that may interfere with focused task execution and evaluation, as described in theories of attention and cognitive load (Kahneman, 1973; Sweller et al., 2011).

Statement Q3 was deliberately worded negatively to counteract the repetition of similar statements and encourage participants to reconsider their views (Guyatt et al., 1999). At the same time, it provided an opportunity for the experiment supervisor to specifically record and document incidental disturbing elements.

Follow-up Questionnaire

Following the study, the respondents completed a second questionnaire on a laptop in a separate room. This took around 8 to 12 minutes. The evaluation scheme was based on a 7-point Likert scale ranging from ‘Strongly Disagree’ to ‘Strongly Agree’. Certain statements required single-choice responses relating to environments or seating concepts.

The questionnaire was divided into four sections. In addition to providing a brief explanation of the Likert scale used, the introduction included general information about the respondents, such as their age, gender, occupation, previous experience with VR, and how often and why they travel by plane.

Part A contained questions about the nine VR settings. Respondents were asked to select the combination of environment and aircraft display that they found most immersive, and to indicate whether the environment influenced their perception of the cabin or the visualisation of the aircraft had a perceptible effect. To improve differentiation and reactivation of visual memory, screenshots of all nine settings were added to the questions. Part B investigated the participants’ presence and was measured using the Presence Questionnaire (PQ) originally developed by Witmer and Singer (Witmer and Singer, 1998). The study employed a revised version of the questionnaire, which was adapted and validated by the Laboratoire de Cyberpsychologie de l’Université du Québec en Outaouais. The instrument measures multiple dimensions of presence, including realism, the possibility to act and examine the environment, interface quality, and self-evaluation of performance. This version preserves the original multidimensional structure, while providing updated wording and improved psychometric robustness for contemporary VR systems (Laboratoire de Cyberpsychologie de l’Université du Québec en Outaouais, 2013).

Participants

The study involved 32 participants from aviation and non-aviation backgrounds (59% male, 41% female; mean age = 33.47 years, SD = 8.98; range = 19–60). The sample size was selected to strike a balance between feasibility and ensuring there was sufficient variability in user characteristics for an exploratory, within-subject VR study.

To support external validity, participants were recruited from a variety of specialist areas, ensuring that the interdisciplinary nature of co-design was reflected in the sample. The specialist cohort comprised 37% researchers, 19% engineers, 16% industrial designers, and 28% other professionals. ‘Other professionals’ refers to individuals who do not specialize in aircraft and cabin development. Having no direct professional connection to the aviation industry, this group deliberately took on the role of non-expert passengers. While we aimed for roughly balanced representation across roles, the sample was not formally stratified by role, experience, age or gender. Therefore, some subgroups may be over- or under-represented.

As prior VR experience could potentially influence the results, participants were asked how frequently they used VR. Of these, 66% had used VR at least once, 16% used it weekly, 9% used it occasionally, and 9% had no prior experience of using VR.

Experimental Procedure

The study was conducted over two consecutive days in the year 2025. The study room was equipped with a large screen on which the VR application was displayed in real time. This allowed the experimental supervisor to observe the movements and interactions of the participant continuously (see Figure 2).



Figure 2: Room setup showing a participant in VR while the experimental supervisor observes and records their responses on a laptop.

Each participant was scheduled to spend around 25 minutes on the VR session and 8 to 12 minutes on the subsequent questionnaire. Before taking

part, all participants signed an information form detailing the potential risks of using VR, as well as a consent form allowing photos to be taken and published, and data to be stored securely and analysed in anonymised form. The 25-minute VR session was divided into the following phases:

- **2-minute** introduction: Setup of the VR headset, familiarisation with navigation within the application's training scene 'Level 0', and a brief overview of the study procedure.
- **23-minute** exploration phase: Participants navigated through nine experimental settings, each consisting of:
 - **1.5 minutes** of free exploration within the respective setting.
 - **1 minute** to answer the three content-related statements about the setting.

To introduce the VR application, Level 0 placed participants in an empty cabin (see Figure 3). Here participants familiarised themselves with the controls and experimented with basic interactions. Once this was complete, the first scenario began.



Figure 3: Participant learning the controls in the empty cabin 'Level 0'.

Participants were instructed to move through the cabin and evaluate the design in terms of usability, comfort, and overall experience. They could move around the cabin freely in each setting. The timing of each phase was controlled using an interval timer on a smartphone. Acoustic signals indicated the transition between the setting exploration phase and the statements round. After completing the application, the participants filled in the follow-up questionnaire themselves.

Data Analysis

The data analysis examined how environmental context and aircraft configuration influence the evaluation of aircraft cabin concepts in VR, while accounting for the within-subjects study design. The three items of the VR session questionnaire (Q1–Q3) were treated as separate dependent variables, representing different aspects of evaluation quality: perceived effectiveness, contextual appropriateness and perceived distraction. To ensure consistent

scale directionality across all dependent variables, responses for the third dependent variable (Q3) were inverted so that higher values reflect a lower number of perceived disruptive factors.

The environment (hangar, airport or sky) and the aircraft configuration (closed, half-open or semi-transparent) were defined as within-subject factors, since all participants experienced all combinations. Prior to inferential analysis, the normality assumption was assessed for each dependent variable using the Shapiro–Wilk test. As the normality assumption was violated for all three variables, a non-parametric variance analysis was required.

To enable the analysis of main effects and interactions in a repeated-measures design with non-normally distributed data, a two-factor repeated-measures Aligned Rank Transformation (ART) analysis of variance (ANOVA) was applied. ART-ANOVA preserves the factorial structure of ANOVA while allowing robust analysis of non-parametric data (Wobbrock et al., 2011). When the omnibus tests were significant, Bonferroni-corrected post hoc pairwise comparisons were conducted on the aligned ranks to control for Type I error inflation.

Responses from the follow-up questionnaire were analysed descriptively. Likert-scale items were summarised using means, standard deviations and confidence intervals, while single-choice items were evaluated using frequency distributions. The aim here was to obtain a more detailed retrospective evaluation of the suitability of the cabin concepts. The full wording of these items can be found in the Appendix.

PQ scores were calculated according to the established scoring procedure and compared descriptively against reference values reported in the literature.

Potential influences of demographic variables (age, gender, professional background) and prior VR experience were examined exploratorily. Since no systematic associations with the primary dependent variables were observed, these factors were not included in further inferential analyses.

Qualitative data from audio recordings of the VR sessions were analysed in a supportive and explanatory manner. Rather than applying a formal coding scheme, participant statements were reviewed to identify recurring themes related to environmental perception, distraction, and evaluation strategies. Selected quotations are reported to contextualise and illustrate the quantitative findings.

RESULTS

The following subsections present the results of the study in three parts: (a) the VR session questionnaire, which assesses immediate evaluation experiences (Q1–Q3); (b) Part A of the follow-up questionnaire, which addresses retrospective preferences and perceived influences; and (c) Part B, which reports presence scores derived from the PQ.

VR Session Questionnaire

The first questionnaire was completed during the VR session and comprised three dependent variables (Q1–Q3). As the Shapiro–Wilk test revealed that all three variables violated the normality assumption (Q1: $W = .86$, $p < .01$; Q2: $W = .89$, $p < .01$; Q3: $W = .83$, $p < .01$), a two-factor repeated-measures ART ANOVA was employed to test the main effects of environment and configuration, as well as their interaction. When the omnibus tests were significant, Bonferroni-corrected post hoc pairwise comparisons were conducted on the aligned ranks for the relevant factor levels. To improve interpretability, the original (untransformed) data's confidence intervals and standard deviations are reported.

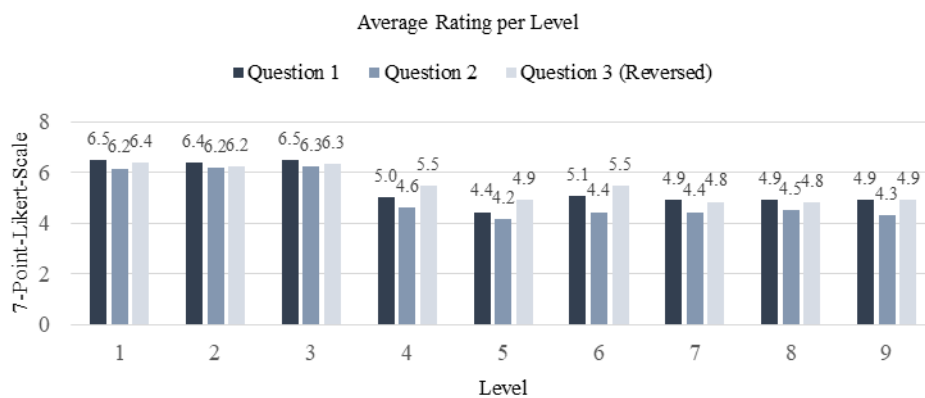


Figure 4: Illustrates the results of the first questionnaire. The average ratings of statements Q1, Q2 and Q3 are shown in the form of grouped bar charts for each setting (Levels 1–9).

To facilitate the categorisation, the statements are labelled under Figure 4. Q3 was coded in reverse for the analysis to ensure direct comparison of the values.

To clarify which level contained which setting, Table 1 provides an overview of the exact composition of each level within the respective settings.

It is notable that levels 1 to 3 with a closed cabin achieve the highest mean values. Conversely, levels 4 to 9 with a half-open or semi-transparent cabin demonstrate a lower overall rating level, yet exhibit comparative homogeneity.

Q1: I was able to evaluate the cabin effectively in this setting

For the dependent variable Q1, the results show clear differences between the cabin and environment variants. Mean ratings on the 7-point Likert scale ranged from 4.44 (Level 5) to 6.50 (Level 3). The closed cabin in the sky (Level 3) achieved the highest approval rating, with a mean value of 6.50. This was closely followed by hangar 6.47 (Level 1) and airport 6.38 (Level 2). The half-open and semi-transparent variants received significantly lower scores,

with average mean values of 4.84 (Levels 4–6) and 4.93 (Levels 7–9). This finding suggests that these forms of presentation are perceived as being less effective than the closed cabin.

Table 1: Composition of VR levels (L1–L9) for aircraft cabin visualisation and environmental contexts.

L1	L2	L3	L4	L5	L6	L7	L8	L9
Closed Cabin			Half-Open Cabin			Semi-Transparent Cabin		
Hangar	Airport	Sky	Hangar	Airport	Sky	Hangar	Airport	Sky

The standard deviations ranged from 0.67 (Level 3) to 1.70 (Level 7), indicating variations in the distribution of ratings across the variants. With standard deviations of 0.91 and 0.67, respectively, Level 1 and Level 3 in particular demonstrated the lowest variance, indicating a high level of agreement between the participants in their assessment. Conversely, higher standard deviations for the half-open and semi-transparent variants are indicative of greater heterogeneity in perception.

It is also important to note that half-open cabin variants (Levels 4–6) exhibit greater variation between the individual environmental representations compared to closed cabins (Levels 1–3). This is evident in the wider range of mean values (0.62 vs. 0.12) and the higher standard deviation of the mean values (SD = 0.34 vs. 0.065).

A comparison of the evaluations of the different environments shows that the sky representation achieved the highest mean values across all three aircraft visualisations: 6.05 (Level 3), 5.06 (Level 6), and 4.94 (Level 9). Conversely, the airport environment consistently received the lowest ratings across all variants and exhibited comparatively high variability in the evaluations, with standard deviations exceeding 1.15 in each case.

An ART repeated-measures ANOVA revealed a small main effect of the environment on task effectiveness, $F(2, 248) = 3.69$, $p = .026$. Bonferroni-adjusted pairwise tests showed ‘Hangar’ > ‘Airport’, $t(248) = 2.651$, $p = .026$, with no differences for ‘Hangar’ vs. ‘Sky’, $t(248) = 0.809$, $p = 1.000$, or ‘Airport’ vs. ‘Sky’, $t(248) = -1.842$, $p = .200$. In contrast, the configuration had a large effect, $F(2, 248) = 83.44$, $p < .001$: ‘Closed’ > ‘Half-open’, $t(248) = 11.963$, $p < .001$, and ‘Closed’ > ‘Semi-transparent’, $t(248) = 10.204$, $p < .001$, with no difference between the two open variants, $t(248) = -1.758$, $p = .240$. The Environment \times Configuration interaction was not significant, $F(4, 248) = 1.23$, $p = .299$.

Q2: The environment and aircraft visualisation felt appropriate for the cabin evaluation.

A similar picture emerges with regard to the second dependent variable Q2. The closed cabin in the sky (Level 3) attained the highest mean score of 6.25, followed by Level 2 (6.19) and Level 1 (6.16). In contrast, the ratings for the half-open and semi-transparent variants were also notably lower in terms of appropriateness, starting at 4.59 (Level 4).

The variability of the ratings was also greater than for the first dependent variable. The standard deviations ranged from 0.95 (Level 3) to 1.90 (Level 6), indicating a slightly higher variance in response behaviour. Nevertheless, the closed cabin remained the most consistent variant, with the lowest standard deviation.

Within the half-open variants, a clear polarisation was observed, reflected in comparatively high standard deviations between 1.68 and 1.90. The airport environment received the lowest rating of all the settings examined, with a mean value of 4.16. In comparison, the half-open hangar environment achieved a mean value of 4.60, ranking fourth overall when compared to the other environments.

An ART-ANOVA showed no main effect of Environment regarding appropriateness, $F(2,248) = 0.45$, $p = .636$, but a strong main effect of Configuration, $F(2,248) = 81.54$, $p < .001$. The Environment \times Configuration interaction was not significant, $F(4,248) = 0.57$, $p = .685$. Bonferroni-corrected pairwise tests indicated that 'Closed' was rated higher than 'Half-open', $t(248) = 11.04$, $p < .001$, and 'Semi-transparent', $t(248) = 11.08$, $p < .001$, while 'Half-open' and 'Semi-transparent' did not differ, $t(248) = 0.05$, $p = 1.00$.

Q3: I noticed elements that made evaluation more difficult.

As with the other statements, the closed cabin variants showed a clear advantage, achieving higher ratings than the half-open and semi-transparent variants.

After inverting the scale, the mean values were between 6.41 (Level 1) and 4.81 (Level 8). The closed variants (Level 1-3) achieved the highest values with scores of 6.41 (Level 1), 6.22 (Level 2) and 6.34 (Level 3), which indicates the lowest perception of disturbing elements during the evaluation. It is striking that, compared to the previous statements, the hangar environment was slightly ahead of the sky environment in this case.

The standard deviations for this statement ranged from 0.95 to 1.90, which is consistent with the values observed for the other statements. Notably, the closed cabin variants showed lower overall dispersion, suggesting a high degree of consistency in participants' perceptions of minor disruptive factors.

In comparison, the half-open and semi-transparent variants exhibited lower mean values, indicating more disruptions, and exhibited greater variance, suggesting greater disagreement and more variable disruptive influences. Level 8 (the semi-transparent cabin with the airport environment) performed worst in this regard.

It is worth noting that the semi-transparent variants performed worse than the half-open variants on this statement compared to the previous ones (Q1 and Q2). This suggests that the semi-transparent representation was perceived as a more disruptive factor.

For the disturbance factors the environment was not significant ($F(2,248) = 2.46$, $p = .088$), whereas the configuration showed a major effect ($F(2,248) = 39.46$, $p < .001$): the closed cabin was perceived as less disturbing than half-open ($t(248) = -7.10$, $p < .001$) and semi-transparent

($t(248) = -8.18, p < .001$), with no difference between half-open and semi-transparent ($t(248) = -1.08, p = .848$); the Environment \times Configuration interaction was non-significant ($F(4,248) = 0.85, p = .497$).

Follow-up Questionnaire

Part A

Consistent with the VR session questionnaire, the sky environment received the highest overall preference in the follow-up questionnaire (56%), followed by the hangar (28%) and the airport (16%). While these results align with earlier evaluations, the difference between the sky and hangar environments was less pronounced in the VR session questionnaire.

The preference for the aircraft representation was even more pronounced. A total of 85% of participants selected the closed configuration as the most effective form of representation, compared to 9% who favored the semi-transparent variant and 6% who preferred the half-open variant.

Participants were also asked to evaluate the different combinations of cabin configuration and environmental context. In addition to assessing immersion and support for spatial perception, they were asked to indicate the combination in which they felt most comfortable and the one that negatively affected their perception. Across all four questions, a consistent pattern emerged, with the closed cabin combined with the sky environment receiving the highest approval ratings for immersion (46.9%), understanding of spatial qualities (40.6%), and perceived comfort (37.5%). Although the closed cabin combined with the hangar environment received low approval ratings for immersion (12.5%), it was rated notably higher for spatial quality (25%) and perceived comfort (22%). Conversely, the transparent cabin combined with the sky and the open cabin combined with the airport environment were most frequently associated with negative effects on perception (25% each).

Subsequent evaluation of the statement “The surrounding environment influenced my judgment of the cabin design.” using a 7-point Likert scale resulted in a mean score of 4.9, with a standard deviation of 1.41. This suggests that participants tended to perceive the environment as having a moderate influence on their evaluation of the cabin design. The relatively large standard deviation indicates variability in individual responses. With a 95% confidence interval of [4.37; 5.38], this assessment clearly exceeds neutrality and remains firmly outside the negative rating range (1–3).

The statement “The aircraft visualisation affected how I perceived the cabin design.” received a mean score of 5.5, with a standard deviation of 1.24. The lower standard deviation compared to the previous item suggests greater agreement among participants regarding this effect. The 95% confidence interval [5.08; 5.98] shows that participants attributed strong influence to aircraft visualisation in shaping their perception of the cabin design.

Part B

Without analysing each item of the PQ in detail at this point, the results nevertheless demonstrate the capacity of the VR application to induce immersion. With a total score of 107.38, the result slightly exceeds Witmer

and Singer's threshold of 104 points, above which an application is considered to exhibit a high level of presence [25]. Above normal values were also observed across all individual categories: Realism (36.53; +7.08), Possibility to Act (23.38; +2.62), Interface Quality (17.00; +1.63), Possibility to Examine (17.66; +2.28) and Self-Evaluation of Performance (12.81; +1.81).

DISCUSSION

"Just like in real life." Addressing RQ1, the results from both questionnaires indicate a clear difference in perceived evaluation quality across virtual settings. Closed cabin variants were consistently preferred for tasks requiring focused attention and minimal distraction. These variants were rated as more effective and appropriate, and were associated with significantly fewer perceived disruptive elements. In contrast, half-open and semi-transparent variants received lower approval ratings, with greater variability in responses indicating a less consistent and more subjective user experience.

Further analysis of the quantitative data showed that the environment did not emerge as a primary driver of evaluation outcomes. Effects were observed solely for task effectiveness (Q1), in which the hangar outperformed the airport. In contrast, cabin configuration was a key factor in all areas. Participants consistently judged the closed cabin to be more effective and appropriate (Q1 & Q2), as well as the least disturbing (Q3), compared to the open variants. Taken together, these results suggest that, in VR-based cabin evaluations, configuration fidelity is paramount, while the design of the surrounding virtual environment can either stabilise or disrupt the evaluation process. Therefore, a closed configuration provides a robust baseline for iterative assessments, whereas open variants should be used selectively and purposefully depending on the evaluation goal.

A likely explanation for the limited influence of the environment is that, of the nine tested settings, three involved the cabin being fully closed. This restricted visibility of the external context, thereby reducing its perceptual relevance during the task-focused evaluation. In addition, as all environments were aviation-related (hangar, airport, sky), they were all plausible within the evaluation context, which may have reduced the perceived differences between them. Under these conditions, subjective preferences appear to have outweighed objective environmental distinctions.

In response to RQ2, three categories of environmental characteristics emerged. Environmental features were considered relevant if they offered coherent contextual cues and perceptual depth information. For instance, outdoor sky views were perceived as supporting immersion, comfort and spatial understanding. They were considered negligible when the cabin configuration obstructed the view of the environment, rendering contextual differences largely irrelevant during focused evaluation. Environmental characteristics were considered limiting if they introduced competing visual salience or perceptual inconsistencies, such as high-detail airport HDRI environment, scaling mismatches or visual discontinuities at open cabin edges. Rendering choices that undermined spatial interpretation, such as

semi-transparency being perceived as a surface texture or the production of confusing shadow artefacts, were also identified as limiting factors.

The follow-up questionnaire revealed that combining a closed cabin with a sky environment enhanced the immersive experience and the perception of the cabin's spatial dimensions, while also improving participants' overall sense of comfort. This effect was particularly evident in reflective judgments after the VR experience, suggesting that environmental context may play a stronger role in post-hoc evaluation than during task-focused interaction. However, the reduced visibility in the closed cabin models limited their impact on cabin perception. As a result, the hangar and airport environments were rated slightly lower in the VR session questionnaire, yet were still considered largely appropriate and effective for closed cabin settings. In contrast, when participants later reflected on the experience, they clearly preferred the outdoor environment.

Additional qualitative data collected during the VR sessions offered a deeper understanding of how different stakeholders perceived the environments. Non-expert participants consistently rated the hangar as their least preferred setting, with 44% reporting feelings of unfamiliarity or discomfort. In contrast, 50% of engineers selected the hangar as their preferred environment. These contrasting preferences highlight that the perceived suitability of a virtual environment is strongly influenced by prior experience and professional background rather than indicating a universally preferable context.

Further differences emerged between the groups of participants in relation to their professional roles. While 83% of engineers reported that the environment was largely irrelevant in closed configurations, 80% of designers emphasised its importance. Designers emphasised the importance of depth perception, which is ideally achieved through window views, and described dynamic environmental cues, such as sound, as crucial for realistically simulating aircraft movement. These diverging views further highlight the need for adaptable VR environments that can be tailored to different stakeholder perspectives and evaluation strategies.

In the airport environment, many participants reported being distracted by external visual stimuli, which negatively impacted their ability to focus on evaluating the cabin. The high level of detail in the HDRI environment, combined with scaling issues in some cases, created what one participant described as "a multitude of opportunities for discovery". These findings suggest that increased visual richness does not necessarily improve the quality of evaluations and may instead undermine them when spatial coherence is disrupted, particularly in open cabin configurations. The frequently criticised visual break-off edge at the open cabin side likely contributed to the lower effectiveness (4.4) and appropriateness (4.2) ratings compared to other settings.

The concept of a semi-transparent wall, which has been shown to encourage creativity in physical work environments, could not be evaluated in the context of a conceptual evaluation (Thoring et al., 2020). The semi-transparency was not consistently perceived in the virtual representation;

it was sometimes interpreted as a static visual texture rather than a spatial feature. Additionally, the resulting shadow effects caused confusion in terms of spatial perception. This suggests that design strategies derived from physical creative environments may not directly translate to virtual evaluation contexts. Nevertheless, this supports the assumption that poorly designed virtual spaces can restrict cognitive engagement and limit creative freedom (Oxman, 2011).

However, the relatively small sample size ($n = 32$) and the specific composition of the sample inevitably limit the generalizability of the results. Although the interdisciplinary composition of the sample supports its contextual relevance, it was not systematically stratified by stakeholder role, level of experience, age or gender ratio. This could affect the transferability of the results to broader populations and contexts. Currently, there is no generally accepted, standardized assessment method for VR-based cabin co-design. The task design, metrics, and rating scales employed in this study represent one approach, but they may emphasize certain aspects of performance and experience over others. Future work should replicate these results using larger, stratified samples and compare alternative assessment protocols to improve robustness and external validity.

Furthermore, the study focused exclusively on the visualisation of an aircraft cabin in three aviation-related environments (hangar, airport, and sky). While this setting is highly relevant to the target application domain, it limits the generalizability of the findings to other design contexts. Different products and industries may require distinct VR evaluation methods, so future research should examine a broader range of use cases and environments. Nevertheless, the underlying insights into how environmental context and configuration shape user evaluations are not confined to identifying an optimal solution, but rather to understanding which contextual characteristics support or hinder evaluation under specific conditions. As such, similar considerations are likely to apply in other domains where VR is used as an evaluation tool, particularly in early-stage design and collaborative assessment scenarios.

CONCLUSION

This study investigated how the design of IVEs influences the evaluation of aircraft cabin concepts, providing foundational insights for future XR-based co-design processes. The results demonstrate that visualising the aircraft cabin itself plays a primary role in shaping spatial perception, whereas the surrounding IVE contributes in a more supportive manner. Rather than being a dominant factor itself, the IVE establishes a cognitively coherent context that supports immersion and stabilises individual evaluation experiences.

In this regard, a closed cabin combined with a sky environment proved to be a robust and widely accepted configuration for evaluation purposes. Although individual preferences varied, participants consistently perceived this context as appropriate and non-distracting, which enabled the reliable assessment of cabin designs.

These findings suggest that effective IVEs for cabin evaluation do not need to maximize realism or complexity, but should instead provide sufficient

spatial depth and contextual coherence while minimizing irrelevant or distracting elements. Achieving this balance supports consistent evaluation across users and provides a meaningful baseline for subsequent investigations of collaborative co-design scenarios.

Future research may build on these findings by examining more complex XR setups and the role of multisensory factors, such as lighting, sound and level of detail. These parameters should be systematically explored to understand their influence on collaborative evaluation and decision-making among diverse stakeholders.

ACKNOWLEDGMENT

We would like to thank all the participants who took part in the study, as well as those who provided useful feedback on earlier versions of this document, and everyone who helped us with this project.

REFERENCES

- Bartl, A., Merz, C., Roth, D., Latoschik, M.E., 2022. The Effects of Avatar and Environment Design on Embodiment, Presence, Activation, and Task Load in a Virtual Reality Exercise Application, in: 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Presented at the 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), IEEE, Singapore, Singapore, pp. 260–269. <https://doi.org/10.1109/ISMAR55827.2022.00041>
- Chan, C.-S., 2011. Design Representation and Perception in Virtual Environments, in: Wang, X., Tsai, J.J.-H. (Eds.), Collaborative Design in Virtual Environments. Springer Netherlands, Dordrecht, pp. 29–40. https://doi.org/10.1007/978-94-007-0605-7_3
- Cornelje, S., Reimer, F., Herzig, J., Lindlar, M., Weiand, P., Nagel, B., 2024. Enhancing The XR+ Method: A Use Case in Rescue Helicopter Cabin Design, in: 2024. Presented at the DLRK, Hamburg.
- Diehl, M., Stroebe, W., 1987. Productivity Loss In Brainstorming Groups: Toward the Solution of a Riddle. *Journal of Personality and Social Psychology* 53(3), 497–509. <https://doi.org/https://doi.org/10.1037/0022-3514.53.3.497>
- Dorta, T., Safin, S., Boudhraa, S., Beaudry Marchand, E., 2019. Co-Designing in Social VR - Process awareness and suitable representations to empower user participation. Presented at the CAADRIA 2019: Intelligent & Informed, Wellington, New Zealand, pp. 141–150. <https://doi.org/10.52842/conf.caadria.2019.2.141>
- El-Jarn, H., Southern, G., 2020. Can co-creation in extended reality technologies facilitate the design process? *JWAM* 12, 191–205. <https://doi.org/10.1108/JWAM-04-2020-0022>
- Froman, R.D., 2014. The ins and outs of self-report response options and scales. *Res Nurs Health* 37, 447–451. <https://doi.org/10.1002/nur.21626>
- Gibson, J.J., 1979. *The Ecological Approach to Visual Perception*. Houghton Mifflin.
- Guyatt, G.H., Cook, D.J., King, D., Norman, G.R., Kane, S.L., Van Ineveld, C., 1999. Effect of the framing of questionnaire items regarding satisfaction with training on residents' responses *Acad. Med.*, 192–194. <https://doi.org/10.1097/00001888-199902000-00018>

- Havig, P., McIntire, J., Geiselman, E., 2011. Virtual reality in a cave: limitations and the need for HMDs?, in: Marasco, P.L., Havig, P.R. (Eds.), . Presented at the SPIE Defense, Security, and Sensing, Orlando, Florida, United States, p. 804107. <https://doi.org/10.1117/12.883855>
- Herzig, J., Reimer, F., Cornelje, S., Biedermann, J., Nagel, B., 2025. Advancing Sustainable Prototyping of Future Aircraft Cabin Designs Through Extended Reality Technologies, in: EASN 2024. Presented at the The 14th EASN International Conference, MDPI, p. 22. <https://doi.org/10.3390/engproc2025090022>
- International Organization for Standardization., 2019. ISO 9241-210:2019 Ergonomics of human–system interaction—Part 210: Human-centred design for interactive systems (No. 210). International Standard, Geneva, Switzerland.
- Kahneman, D., 1973. Attention and effort, Prentice-Hall series in experimental psychology. Prentice-Hall, Inc, Englewood Cliffs, New Jersey.
- Khairul, Purnama, I., Prananta, A.W., Megahati.S, R.R.P., Agustin, R., 2025. Advantages and Disadvantages of Virtual Reality in Science Learning Systems in the 21st Century: A Review. *jppipa, pendidikan ipa, fisika, biologi, kimia* 11, 35–41. <https://doi.org/10.29303/jppipa.v11i2.10214>
- Kozhevnikov, M., Garcia, A., 2011. Visual-Spatial Learning and Training in Collaborative Design in Virtual Environments, in: Wang, X., Tsai, J.J.-H. (Eds.), Collaborative Design in Virtual Environments. Springer Netherlands, Dordrecht, pp. 17–26. https://doi.org/10.1007/978-94-007-0605-7_2
- Krug, S., 2013. Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability, 3rd ed. New Riders.
- Kusmaryono, I., Wijayanti, D., Maharani, H.R., 2022. Number of Response Options, Reliability, Validity, and Potential Bias in the Use of the Likert Scale Education and Social Science Research: A Literature Review. *INT J EDUC METHODOL* 8, 625–637. <https://doi.org/10.12973/ijem.8.4.625>
- Laboratoire de Cyberpsychologie de l'Université du Québec en Outaouais, (UQO), 2013. Questionnaire Sur L'État De Présence. Québec.
- Maher, M.L., 2011. Designers and Collaborative Virtual Environments, in: Wang, X., Tsai, J.J.-H. (Eds.), Collaborative Design in Virtual Environments. Springer Netherlands, Dordrecht, pp. 3–15. https://doi.org/10.1007/978-94-007-0605-7_1
- Man, D., Olchawa, R., 2018. The Possibilities of Using BCI Technology in Biomedical Engineering, in: Hunek, W.P., Paszkiel, S. (Eds.), Biomedical Engineering and Neuroscience, Advances in Intelligent Systems and Computing. Springer International Publishing, Cham, pp. 30–37. https://doi.org/10.1007/978-3-319-75025-5_4
- Moerland-Masic, I., Reimer, F., Bock, T.M., Meller, F., Nagel, B., 2022. Application of VR technology in the aircraft cabin design process. *CEAS Aeronaut J* 13, 127–136. <https://doi.org/10.1007/s13272-021-00559-x>
- Neo, J.R.J., Won, A.S., Shepley, M.M., 2021. Designing Immersive Virtual Environments for Human Behavior Research. *Front. Virtual Real.* 2, 603750. <https://doi.org/10.3389/frvir.2021.603750>
- Nik Ahmad, N.A., Hasni, N.S., 2021. ISO 9241-11 and SUS Measurement for Usability Assessment of Dropshipping Sales Management Application, in: 2021 10th International Conference on Software and Computer Applications. Presented at the ICSCA 2021: 2021 10th International Conference on Software and Computer Applications, ACM, Kuala Lumpur Malaysia, pp. 70–74. <https://doi.org/10.1145/3457784.3457794>

- Oxman, R., 2011. Design Paradigms for the Enhancement of Presence in Virtual Environments, in: Wang, X., Tsai, J.J.-H. (Eds.), Collaborative Design in Virtual Environments. Springer Netherlands, Dordrecht, pp. 41–49. https://doi.org/10.1007/978-94-007-0605-7_4
- Paulus, P.B., Brown, V.R., 2007. Toward More Creative and Innovative Group Idea Generation: A Cognitive-Social-Motivational Perspective of Brainstorming. *Social & Personality Psych* 1, 248–265. <https://doi.org/10.1111/j.1751-9004.2007.00006.x>
- Pilcher, N., Cortazzi, M., 2024. “Qualitative” and “quantitative” methods and approaches across subject fields: implications for research values, assumptions, and practices. *Qual Quant* 58, 2357–2387. <https://doi.org/10.1007/s11135-023-01734-4>
- Porqueddu, E., 2025. Designing the Complex City: A Systemic Approach to Spatial Design, 1st ed. Routledge, New York. <https://doi.org/10.4324/9781003495659>
- Prithul, A., Adhanom, I.B., Folmer, E., 2021. Teleportation in Virtual Reality; A Mini-Review. *Front. Virtual Real.* 2, 730792. <https://doi.org/10.3389/frvir.2021.730792>
- Rauschnabel, P.A., Babin, B.J., Tom Dieck, M.C., Krey, N., Jung, T., 2022. What is augmented reality marketing? Its definition, complexity, and future. *Journal of Business Research* 142, 1140–1150. <https://doi.org/10.1016/j.jbusres.2021.12.084>
- Reimer, F., Fuchs, M., Herzig, J., Cornelje, S., Biedermann, J., Nagel, B., 2024. Investigating the Flexible Hydrogen Aircraft System Interface Design using the XR Co-Design Approach ICAS Conference Proceedings.
- Sanders, E.B.-N., Stappers, P.J., 2019. Convivial Toolbox: Generative Research for the Front End of Design, 3rd ed. BIS Publishers, Amsterdam.
- Sanders, E.B.-N., Stappers, P.J., 2008. Co-creation and the new landscapes of design. *CoDesign* 4, 5–18. <https://doi.org/10.1080/15710880701875068>
- Santhosh, S., De Crescenzo, F., 2023. A Mixed Reality Application for Collaborative and Interactive Design Review and Usability Studies, in: Gerbino, S., Lanzotti, A., Martorelli, M., Mirálbes Buil, R., Rizzi, C., Roucoules, L. (Eds.), Advances on Mechanics, Design Engineering and Manufacturing IV, Lecture Notes in Mechanical Engineering. Springer International Publishing, Cham, pp. 1505–1515. https://doi.org/10.1007/978-3-031-15928-2_131
- Schmohl, T. (Ed.), 2021. Situiertes Lernen im Studium: Didaktische Konzepte und Fallbeispiele einer erfahrungsbasierten Hochschullehre. wbv Publikation, Bielefeld. <https://doi.org/10.3278/6004731w>
- Skarbez, R., Smith, M., Whitton, M.C., 2021. Revisiting Milgram and Kishino’s Reality-Virtuality Continuum. *Front. Virtual Real.* 2, 647997. <https://doi.org/10.3389/frvir.2021.647997>
- Sweller, J., Ayres, P., Kalyuga, S., 2011. Cognitive Load Theory. Springer New York, New York, NY. <https://doi.org/10.1007/978-1-4419-8126-4>
- Thoring, K., Mueller, R.M., Desmet, P., Badke-Schaub, P., 2020. Spatial design factors associated with creative work: a systematic literature review. *AIEDAM* 34, 300–314. <https://doi.org/10.1017/S0890060420000232>
- Tullis, T., Albert, B., 2013. Measuring the User Experience, 2nd ed. Elsevier.
- Visser, F.S., Stappers, P.J., Van Der Lugt, R., Sanders, E.B.-N., 2005. Contextmapping: experiences from practice. *CoDesign* 1, 119–149. <https://doi.org/10.1080/15710880500135987>
- Wang, X., Wang, R., 2011. Co-presence in Mixed Reality-Mediated Collaborative Design Space, in: Wang, X., Tsai, J.J.-H. (Eds.), Collaborative Design in Virtual Environments. Springer Netherlands, Dordrecht, pp. 51–64. https://doi.org/10.1007/978-94-007-0605-7_5

- Waterlander, W.E., Jiang, Y., Steenhuis, I.H.M., Ni Mhurchu, C., 2015. Using a 3D Virtual Supermarket to Measure Food Purchase Behavior: A Validation Study. *J Med Internet Res* 17, e107. <https://doi.org/10.2196/jmir.3774>
- Wise, J.A., Hopkin, V.D., Garland, D.J. (Eds.), 2010. *Handbook of aviation human factors*, 2nd ed. ed, Human factors in transportation. CRC Press, Boca Raton. <https://doi.org/https://doi.org/10.1201/b10401>
- Witmer, B.G., Singer, M.J., 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence* 7, 225–240. <https://doi.org/10.1162/105474698565686>
- Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J., 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Presented at the CHI '11: CHI Conference on Human Factors in Computing Systems, ACM, Vancouver BC Canada, pp. 143–146. <https://doi.org/10.1145/1978942.1978963>

APPENDIX

Follow-up Questionnaire Items “PART A”

1. Which combination (environment + aircraft) was the best to immerse in the cabin?
2. Which combination (environment + aircraft) helped you best understand the cabin’s spatial qualities?
3. In which combination (environment + aircraft) did you feel most confident in evaluating the cabin?
4. Did any combination (environment + aircraft) negatively affect your perception of the cabin?
5. The surrounding environment influenced my judgment of the cabin design.
6. The aircraft visualisation affected how I perceived the cabin design.
7. The virtual cabin experience felt immersive.
8. The cabin appeared realistic across the different environments and aircraft visualisations.