Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 841–851
https://doi.org/10.54941/ahfe1008316

Exploring Usability and User Perception in an Immersive Robotics Learning Environment

Diogo Silva[1,2], Eduardo Santos[1], Miguel Nunes[1], Carlos Costa[3], Francisco Silva[3,4], Arabela Fabre[3], Carolina Pimenta[1], Nuno Sousa[1,4], and Iara Margolis[1,5]

[1]CCG/ZGDV Institute, Guimarães, 4800, Portugal
[2]Universidade do Minho, Braga, Portugal
[3]CTCV – Technological Centre for Ceramics and Glass Coimbra, Portugal
[4]Centro Algoritmi, Universidade do Minho, Guimarães, 4800, Portugal
[5]Lab2PT (Landscape, Heritage and Territory Laboratory), Universidade do Minho, Guimarães, 4800, Portugal

ABSTRACT

This paper presents an early-stage formative evaluation of the usability and user experience of an immersive Virtual Reality (VR) robotics learning environment developed for the Meta Quest 3 headset. A convergent mixed-methods approach combined concurrent think-aloud, task-based observation, semi-structured post-task interviews, and four standardized questionnaires: the System Usability Scale (SUS), the Virtual Reality System Usability Questionnaire (VRSUQ), the Virtual Embodiment Questionnaire (VEQ), and the Simulator Sickness Questionnaire (SSQ). Fourteen participants with expertise in robotics and/or XR completed individual sessions in a controlled environment. SUS ($M = 79.46$, $SD = 11.69$) and VRSUQ ($M = 79.17$, $SD = 11.20$) scores indicated good perceived usability. However, VRSUQ Effectiveness scores were significantly lower than Efficiency and Satisfaction, revealing weaknesses in feedback clarity and error prevention. The qualitative findings identified recurring problems involving interaction feedback, language and speech consistency, spatial orientation, control mechanisms and physical discomfort. The findings informed seven priority redesign recommendations. Given the small, purposively selected expert sample, the results should be interpreted as context-specific pilot evidence supporting iterative development rather than as generalizable evidence of effectiveness for industrial learners.

Keywords: Virtual reality, XR education, Usability, User experience, Human-computer interaction, Robotics education, Virtual embodiment, Immersive learning

INTRODUCTION

The accelerating adoption of robotics in industrial settings, particularly in traditional manufacturing sectors such as ceramics, glass, and stone, is generating a growing demand for effective and accessible operator training solutions. Conventional training approaches, dependent on physical equipment and on-site instruction, impose significant logistical and economic

Received September 17, 2026; Revised October 20, 2026; Accepted October 27, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

constraints, particularly for small and medium enterprises seeking to upskill their workforce (Perez et al., 2019). Virtual reality (VR), through immersive and interactive environments, has emerged as a compelling alternative: immersive simulators can support experiential learning about robotic systems while substantially reducing the costs and risks associated with hands-on training (Verner et al., 2022). The substitution of physical robots with virtual counterparts can make robotics education more equitable and scalable across industrial contexts (Perez et al., 2019).

Research supports the potential of immersive VR for developing practical skills and increasing engagement in science and engineering education (Yang et al., 2024; Yu, 2022). XR-based robotics training frameworks have also been explored in this context (Mulero-Perez et al., 2025). In industrial robotics training, virtual environments can provide opportunities to explore spatial relationships, operational sequences, and technical concepts without relying continuously on physical equipment. However, technological immersion alone does not ensure an effective learning experience. The educational value of VR also depends on interaction quality, feedback, spatial orientation, physical comfort, and the extent to which the system supports users in completing and understanding the intended activities.

These issues make formative usability evaluation particularly important during the early stages of VR development. Generic usability measures can provide standardized benchmarks, but they may not fully capture VR-specific aspects such as embodiment, interaction fluency, fatigue, and simulator sickness (Bareïšytė et al., 2024; Kim and Rhiu, 2024). Combining standardized instruments with observation and participants' verbalized experiences can therefore provide a more detailed basis for iterative redesign.

The present study reports an early-stage formative evaluation of the first functional iteration of an immersive VR robotics learning environment. Participants with expertise in robotics and/or XR were intentionally recruited to identify high-impact interaction, content, and system-level problems before subsequent evaluation with the intended industrial end-users. The study does not assess learning effectiveness or workforce training outcomes; rather, it examines perceived usability and user experience during interaction with the prototype.

Using concurrent think-aloud, task-based observation, post-task interviews, and standardized questionnaires, the study aimed to identify usability barriers, examine participants' perceptions of the interaction, and formulate evidence-based recommendations for the next design iteration. Accordingly, the study addressed the following research question: How do expert users perceive the usability and user experience of an early-stage immersive VR robotics learning environment, and which interaction factors should inform its subsequent redesign?

METHOD

The study adopted a convergent mixed-methods formative evaluation design, integrating qualitative and quantitative evidence collected during a single individual session. Observational and verbal data were used to identify interaction problems and understand participants' perceptions, while

standardized questionnaires provided descriptive benchmarks for perceived usability, embodiment, and simulator sickness. Sessions were conducted individually in a controlled environment using a Meta Quest 3 headset.

Instruments

The usability evaluation combined qualitative and quantitative data collection instruments. During interaction, participants were asked to verbalize their thoughts using a concurrent think-aloud protocol. Neutral prompts, such as "What are you thinking right now?", were used during prolonged silences. Participants could skip a task after 90 seconds of complete blockage, without receiving procedural guidance from the moderator.

Interaction was structured around a 15-task usability guide covering the principal flows available in M1. Representative operations included selecting and grabbing virtual objects, navigating between the tutorial, lobby, and thematic floors, using the elevator interface, completing diagnostic questions, interacting with information displays, and attempting to operate a virtual robotic arm through a joystick. Tasks were presented orally and sequentially without revealing subsequent steps.

The post-task semi-structured interview addressed the overall experience, mental-model clarity, navigation, interaction feedback, visual presentation, cognitive demands, moments of difficulty, positive aspects, and improvement suggestions. The guide provided a consistent structure while allowing the moderator to explore issues that emerged during each session.

Four standardized questionnaires were also administered. The System Usability Scale (SUS) (Brooke, 1996) is a widely used instrument for assessing perceived system usability through a global score ranging from 0 to 100. Scores were interpreted using the adjective rating framework proposed by Bangor et al. (2009). Approximate interpretative ranges were considered as follows: below 51 (unacceptable), 51–62 (poor), 63–72 (acceptable), 73–85 (good), 86–90 (excellent), and above 90 (best imaginable). These ranges were treated as descriptive guidelines rather than strict categorical thresholds, and SUS scores were interpreted without prior rounding.

The Virtual Reality System Usability Questionnaire (VRSUQ) (Kim and Rhiu, 2024) is a nine-item VR-specific instrument comprising three dimensions: Effectiveness, Efficiency, and Satisfaction. Items are rated on a five-point scale, and the overall score can be converted to a 0–100 scale. The instrument was included to complement the generic SUS score with VR-specific information concerning feedback, interaction, error handling, learnability, and physical and emotional comfort.

The Virtual Embodiment Questionnaire (VEQ) (Roth and Latoschik, 2020) comprises 12 items rated on a seven-point scale and assesses ownership, agency, and perceived body change. It was included exploratorily to establish a baseline for participants' perceived control and embodiment during interaction with the VR environment.

The Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993) assesses 16 symptoms on a four-point scale and provides Nausea, Oculomotor, Disorientation, and Total Severity scores. It was administered immediately after VR exposure to document physical discomfort associated with the session.

VR Learning Environment

The complete VR environment comprises three modules that progressively address introductory concepts, structured technical content, and practical simulation. The present study evaluated only Module 1 (M1), which introduces robot classifications, industrial automation, collaborative robotics, actuators, and sensors. Content is delivered through a virtual robotic tutor, information displays, interactive demonstrations, diagnostic questions, and object-manipulation activities. Humour and informal communication are used to support engagement during learners' first contact with the robotics content.

Procedure

Each session lasted approximately one hour and comprised four phases. First, participants received information about the study, provided informed consent, completed a sociodemographic questionnaire and received standardized instructions on the Meta Quest 3 controllers. Second, participants completed the 15 interaction tasks while using the headset. The headset view was cast to the investigator's laptop to support observation and contemporaneous field-note recording. Third, participants completed the post-exposure SSQ, SUS, VRSUQ, and VEQ. Finally, a semi-structured post-task interview was conducted.

Participants

Fourteen participants were purposively recruited for this early-stage formative evaluation. Their areas of expertise included robotics (n = 6), XR and related technology development (n = 5), both robotics and XR (n = 2), and education, skills development and vocational training (n = 1). The sample included 10 men and 4 women aged between 23 and 48 years (M = 30.5, SD = 8.7). Seven participants held a bachelor's degree, six held a master's degree, and one held a doctoral degree. Regarding prior VR experience, one participant reported no familiarity, six reported limited familiarity, four reported moderate familiarity, and three reported extensive familiarity.

Participants with relevant domain experience were intentionally selected to provide informed usability feedback during the early stages of the iterative design process. Their involvement supported the identification of high-impact interaction, technical, and content-related issues before subsequent evaluation with novice learners and intended industrial end-users. This sampling strategy was appropriate for formative redesign but does not support generalization to the broader industrial workforce. All participants provided informed consent prior to the study.

Data Analysis

Qualitative findings were derived from contemporaneous written field notes produced during the think-aloud sessions, task observation, and post-task interviews. One researcher inductively coded recurring critical incidents and participant comments, grouped them into design-relevant usability categories, and classified their severity as critical, serious, moderate, or an

opportunity for improvement. Critical problems prevented task completion; serious problems caused recurrent errors or confusion; moderate problems produced recoverable hesitation; and opportunities represented non-blocking improvements. Because complete audio recordings and verbatim transcripts were not available for all sessions, the qualitative findings should be interpreted as formative observational evidence rather than as an exhaustive thematic account of participants' experiences.

Quantitative analysis comprised descriptive statistics, one-sample comparisons of the VEQ dimensions against the neutral midpoint of 4, and a repeated-measures ANOVA comparing the VRSUQ subscales. Distributional assumptions were examined using the Shapiro–Wilk test, and Wilcoxon signed-rank tests were applied when parametric assumptions were not met. Post hoc pairwise comparisons between VRSUQ subscales were conducted using Bonferroni-adjusted p-values, while effect sizes for Wilcoxon signed-rank tests were calculated using the matched-pairs rank-biserial correlation. Given the small sample, inferential findings were interpreted cautiously and reported with effect sizes. Quantitative results were used primarily for descriptive benchmarking and triangulation with the observational evidence. SUS scores were calculated according to Brooke (1996) and interpreted using the adjective-based ranges proposed by Bangor et al. (2009). VRSUQ scores and subscales were calculated according to Kim and Rhiu (2024), VEQ dimensions according to Roth and Latoschik (2020), and post-exposure SSQ subscales and Total Severity scores using the weighted procedure proposed by Kennedy et al. (1993).

RESULTS

Quantitative Assessment

All 14 participants completed the four questionnaires. The mean SUS score was 79.46 (SD = 11.69, Mdn = 75), corresponding to "Good" perceived usability according to the adjective-based interpretation proposed by Bangor et al. (2009). The mean VRSUQ score was 79.17 (SD = 11.20). Given the small purposive sample, these scores are interpreted as descriptive indicators of perceived usability rather than as population-level estimates.

VRSUQ subscale scores differed significantly, $F(2, 26) = 6.13$, $p = 0.007$, $\eta^2 = 0.32$. Effectiveness ($M = 3.79$, $SD = 0.56$) was lower than Efficiency ($M = 4.33$, $SD = 0.59$, $p = 0.020$) and Satisfaction ($M = 4.38$, $SD = 0.67$, $p = 0.016$). This pattern suggests that feedback clarity and error prevention were comparatively weaker than perceived learnability, efficiency, and comfort.

No VEQ dimension differed significantly from the neutral midpoint of 4: Ownership, $t(13) = 0.21$, $p = 0.839$, $d = 0.06$; Agency, $W = 80.50$, $p = 0.083$, $r = 0.53$; and Change, $t(13) = -1.27$, $p = 0.226$, $d = -0.34$. Agency showed the strongest positive tendency, although the result did not reach statistical significance. Overall, the environment did not elicit a strong perception of virtual embodiment.

Post-exposure SSQ Total Severity scores were strongly right-skewed ($M = 27.52$, $SD = 42.02$, $Mdn = 13.09$), with scores ranging from 0 to 145.86. Most participants reported no severe individual symptoms;

however, three participants presented elevated Total Severity scores. P11 showed the highest score and reported locomotion-related discomfort during interaction. These results indicate substantial individual variation and support the inclusion of configurable locomotion speed and rotation options in subsequent iterations.

Usability Findings: Qualitative Analysis

The analysis identified five recurring usability problem categories across the 14 sessions: insufficient feedback, tutor language and speech inconsistencies, navigation disorientation, control issues and discomfort and motion sickness. Table 1 summarizes the frequency and severity of these problems and the corresponding redesign priorities.

Insufficient Feedback [P01, P02, P03, P04, P05, P06, P07, P08, P10, P11, P12, P13, P14]. This category concerns the lack of feedback the system exhibited when users interacted with its surroundings: "I don't know if I'm selecting correctly, but there's no feedback" (P11). The most pervasive issue occurred when participants consistently failed to perceive whether their responses in the diagnostic assessment floor were registered. The white selection indicator was repeatedly questioned: "The ball that was supposed to turn green doesn't change. The interaction method is easy, what's missing is the feedback" (P11). Furthermore, some users mentioned the complete absence of auditory feedback: "Sound feedback would help a lot." (P10). This finding aligns with the lower VRSUQ Effectiveness scores and supports redesigning selection states with multimodal feedback (colour change, animation, sound).

Table 1: Usability problem categories, severity, frequency, and priority redesign recommendations.

	Category	Severity	Sessions (N = 14)	Priority Recommendation
1	Insufficient Feedback	Serious–Critical	13	Multimodal feedback redesign
2a	Tutor Language Inconsistency	Serious	6	Enforce language consistency
2b	Tutor Speech Disruption	Moderate	8	Correct the dialogue tree architecture
3	Navigation Disorientation	Moderate–Serious	13	Wayfinding + floor entry messages
4a	Select vs. Grab (tutorial)	Moderate	10	Improve controller instructions
4b	Robotic Arm Joystick	Critical	10	Complete interaction redesign
5	Discomfort / Motion Sickness	Moderate–Serious	6	Locomotion speed + rotation personalization options

Severity scale: Critical = task completion blocked; Serious = recurrent confusion or errors; Moderate = recoverable hesitation; Opportunity = non-blocking improvement.

Tutor Language and Speech Inconsistencies [P01, P02, P03, P05, P06, P07, P08, P09, P10, P11, P13, P14]. This category concerns issues related

to the Tutor's spoken delivery during sessions. Two major problems were identified. (a) Tutor language inconsistency: the robotic tutor switched between Portuguese and English during sessions, including mid-instruction: "Is it normal for him to switch languages? He started in Portuguese and is now speaking in English" (P09). (b) Tutor speech disruption: tutor speech was frequently cut off before completion, disrupting the instructional flow at critical moments: "He stopped talking. I don't know if he was supposed to." (P13).

Navigation Disorientation [P01, P02, P03, P04, P05, P06, P07, P08, P09, P10, P11, P13, P14]. This category concerns the difficulty users had when trying to navigate the world and choosing their next step. Transitions between the tutorial, lobby, and themed floors were a recurrent source of confusion. The elevator panel was particularly problematic, with multiple participants unable to locate the diagnostic assessment floor: "I don't see the diagnostic assessment floor" (P09). Many users mentioned that the absence of a welcome message with orientation instructions, upon entering each floor, further impeded spatial orientation: "There should be some indication, or the tutor should say what's going to happen" (P14).

Control Issues [P01, P02, P03, P04, P08, P09, P10, P11, P12, P13]. This category concerns the issues users had when manipulating the controls. Two major problems were identified. (a) Select vs. grab confusion in the tutorial: most participants resolved quickly. (b) Robotic arm joystick interaction: participants were unable to operate the interactive arm despite on-screen instructions: "I'm moving the joystick and nothing is happening" (P09), constituting a complete interaction blockage in several sessions.

Discomfort and Motion Sickness [P03, P05, P06, P08, P10, P11]. This category concerns users' discomfort with in-game movement. Locomotion speed and the snap rotation feature (body rotation with joystick worked in steps, not in a smooth natural way) were identified as sources of discomfort and interaction-related improvement suggestions, respectively: "The rotation should be more continuous" (P13); "The movement speed is quite fast, I have to make an effort to walk slowly" (P10). The elevator animation was identified as another legitimate trigger: "I think the elevator movement causes vertigo, and moving with the joystick can also cause discomfort" (P11). These findings are corroborated by the SSQ data, most notably P11's severe post-exposure score.

Several sub-problems were identified within these categories. The absence of a playback lock on audio reproduction controls allowed users to trigger simultaneous instances of the same dialogue, resulting in audio overlap: "I'm hearing two things at the same time" (P14). The tutor's speech rate and voice were also criticized, with some participants noting that the pace was excessive and the synthesized quality of the voice detracted from engagement: "The robot speaks a bit too fast. People have different attention levels. I feel that if I were in a different mental state, it would be too much information" (P10); "The Tutor's voice should be less monotonous." (P06). In the diagnostic assessment floor, participants noted that incorrect answers were visually revealed following task completion, yet correct answers were never displayed, limiting the potential for self-correction and consolidation of knowledge: "It doesn't show the correct answer here" (P13). Finally, several

participants suggested increasing the proportion of interactive activities relative to the tutor's expository monologues, in order to better balance instructional delivery with active engagement: "There's too much talking from the robot. I want to do more activities" (P03); "It would be good to have questions to answer, or something more dynamic" (P02).

Positive Aspects. A balanced analysis of the qualitative data reveals that, alongside the usability problems identified, a substantial number of design elements were consistently praised across sessions, suggesting a solid foundation upon which future iterations can build. The robotic tutor's humorous and engaging communication style was among the most frequently cited strengths, perceiving humor as making instructional content more engaging and less monotonous: "The jokes help, they make the theory less boring" (P07). The tutorial and onboarding sequence was widely recognized as well-structured and pedagogically effective, with participants describing it as interactive, sequentially clear, and instrumental in establishing the conceptual and motor foundations needed for subsequent activities: "You understand what the game's objective is. Being able to test the controls is a plus" (P03). The stand-based spatial layout of the floors was perceived as intuitive and non-fatiguing, facilitating navigation and providing a clear sense of task progression: "Good to separate the robot types into stands" (P09). Controller mechanics were broadly described as easy to learn and recall, with the haptic proximity feedback singled out as a meaningful and appreciated affordance: "I liked the haptic feedback when you get close to a button" (P05). The visual design of the environment was noted for its ability to focus participant attention on instructional content without inducing distraction, with colour-coded elements and spatial organization contributing positively to cognitive orientation: "The environment is not boring, it's not garish to the point of distracting from learning. The more vivid colours being on the robots helps keep attention on what matters." (P09). The elevator-based modular structure was valued as a meaningful metaphor for learning progression, creating a coherent navigational logic between programmed stages: "It gives the sense of being levels, it makes the learning more gamified." (P12). The content monitor presenting written information alongside the tutor's speech was identified as supportive of comprehension, particularly for participants who preferred a multimodal presentation of content: "Having the monitors, they should perhaps be in all floors." (P07). Finally, several participants reflected that the immersive VR format rendered the learning of robotics concepts more engaging and memorable compared to traditional didactic methods, noting that physical presence and interactivity created an experiential quality absent from conventional instructional approaches: "Well-designed system, the environment is not boring, the information was well delivered. A more intuitive way to learn." (P09).

DISCUSSION

The quantitative and qualitative results present consistent internal coherence. Global usability scores (SUS and VRSUQ) classify the system as "Good", indicating that users can interact with the system at an acceptable level

of perceived usability. However, VRSUQ subscale analysis and qualitative data reveal that this global assessment conceals specific and relevant problems. The Effectiveness subscale, most directly linked to feedback quality, compatibility, and error prevention, was significantly lower than the others, converging with the insufficient feedback category identified qualitatively as the most frequent and severe problem. This convergence reinforces the validity of the findings and underscores that the highest-impact improvements must focus on these dimensions. These findings are consistent with broader VR usability research: Bareïšytė et al. (2024) highlight that feedback clarity and system responsiveness are among the most frequently reported usability issues in immersive environments, while Kim and Rhiu (2024) note that Effectiveness is typically the weakest dimension across VR systems evaluated with the VRSUQ. Additionally, the consistently positive reception of the tutor's personality and the stand-based layout suggest that the system's engagement strategy and spatial organization represent design strengths that subsequent iterations should preserve and extend.

These findings should not be interpreted as evidence that the environment improves robotics learning or is ready for workforce upskilling. Participants were purposively selected based on their experience in robotics and/or XR, rather than as representatives of the intended industrial learner population. Their role was to identify high-impact technical, interaction, and content-related issues before evaluation with novice learners and industrial end-users. The results therefore constitute context-specific pilot evidence intended to support redesign, rather than generalizable conclusions about learning effectiveness or workforce adoption.

Based on the convergent analysis, seven priority redesign recommendations are proposed: (1) introduce multimodal feedback for selection interactions, combining colour change (white → green), button animation, and auditory confirmation; (2) enforce language consistency throughout the session, ensuring the tutor remains in the user's selected language from start to finish; (3) correct the dialogue tree architecture to ensure proper speech flow; (4) improve spatial wayfinding with tutor-guided transitions between spaces and welcome messages at each floor entry; (5) improve controller instructions in the tutorial; (6) completely redesign the robotic arm activation flow with explicit controller indicators, a visible activation state, and a dedicated micro-tutorial; and (7) provide locomotion speed adjustment and continuous rotation as user-configurable accessibility options.

The VEQ results did not indicate strong virtual embodiment. This is not necessarily problematic for the present system, which relies primarily on controller-mediated interaction rather than avatar ownership or full-body representation. The tendency observed for Agency may nevertheless justify further investigation after the control mechanisms have been redesigned.

The SSQ results showed considerable individual variation. Although severe individual symptom ratings were uncommon, three participants presented elevated Total Severity scores, and locomotion-related discomfort was also identified during observation. These findings support configurable movement options and more systematic monitoring of simulator sickness in subsequent evaluations.

Several limitations must be considered. First, the small purposive sample limits statistical power and does not represent novice learners or the intended industrial workforce. Second, the qualitative analysis relied on contemporaneous field notes produced by one researcher because complete audio recordings and verbatim transcripts were unavailable. Consequently, the qualitative categories should be interpreted as formative usability evidence rather than as an exhaustive thematic representation of participants' experiences. Third, the study evaluated only the introductory module and did not assess learning gains, retention, transfer, or workplace performance. Future studies should evaluate the redesigned system with larger and more diverse samples, including industrial end-users, and incorporate learning and task-performance measures.

CONCLUSION

This study conducted an early-stage formative evaluation of an immersive VR robotics learning environment using task observation, concurrent think-aloud, post-task interviews, and standardized questionnaires. Global SUS and VRSUQ scores indicated good perceived usability, but the combined evidence revealed recurring problems involving interaction feedback, language and speech consistency, spatial orientation, control mechanisms and physical discomfort.

The study provides a context-specific empirical and methodological contribution to the evaluation of immersive learning environments. Empirically, it shows that favourable global usability scores can coexist with recurrent and, in some cases, task-blocking interaction problems in an early-stage VR robotics training prototype. Methodologically, the convergence of VR-specific questionnaire dimensions with think-aloud and observational evidence illustrates how global usability scores can be diagnostically unpacked to identify problems involving feedback, spatial orientation, control mechanisms, tutor behaviour, and locomotion. This combined evidence was translated into prioritized redesign requirements for the subsequent development of the environment.

Because the study involved only 14 purposively selected participants with relevant domain experience and relied on written observational notes rather than complete audio transcripts, its findings are descriptive, context-specific, and not generalizable to industrial learners. The study does not establish learning effectiveness or readiness for workforce deployment. Future research should evaluate the redesigned environment with larger and more diverse samples, including intended industrial end-users, and examine task performance, learning, retention, and simulator sickness over repeated use.

ACKNOWLEDGMENT

This work has been supported by the European Union under the NextGenerationEU, through a grant of the Portugal's Recovery and Resilience Plan (RRP) Partnership Agreement, within the scope of the project ECP – Ecocerâmica e Cristal de Portugal (Project ref. nr. 76 - C644916391-00000029).

REFERENCES

Bangor, A., Kortum, P., and Miller, J. (2009) "Determining what individual SUS scores mean: Adding an adjective rating scale", Journal of Usability Studies, 4(3), pp. 114–123.

Bareïšytė, L., Slatman, S., Austin, J., Rosema, M., van Sintemaartensdijk, I., Watson, S., and Bode, C. (2024) "Questionnaires for evaluating virtual reality: A systematic scoping review", Computers in Human Behavior Reports, 16, Article 100505.

Brooke, J. (1996) "SUS: A 'quick and dirty' usability scale", in: Jordan, P. W., Thomas, B., Weerdmeester, B. A., and McClelland, A. L. (eds.), Usability Evaluation in Industry. London: Taylor & Francis, pp. 189–194.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993) "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness", International Journal of Aviation Psychology, 3(3), pp. 203–220.

Kim, Y. M., and Rhiu, I. (2024) "Development of a virtual reality system usability questionnaire (VRSUQ)", Applied Ergonomics, 119, Article 104319.

Mulero-Perez, D., Zambrano-Serrano, B., Ruiz Zuniga, E., Fernandez-Vega, M.-A., and Garcia-Rodriguez, J. (2025) "Enhancing robotics education through XR simulation: Insights from the X-RAPT training framework", Applied Sciences, 15(18), Article 10020.

Perez, L., Diez, E., Usamentiaga, R., and Garcia, D. F. (2019) "Industrial robot control and operator training using virtual reality interfaces", Computers in Industry, 109, pp. 114–120.

Roth, D., and Latoschik, M. E. (2020) "Construction of the Virtual Embodiment Questionnaire (VEQ)", IEEE Transactions on Visualization and Computer Graphics, 26(12), pp. 3546–3556.

Verner, I., Cuperman, D., Perez-Villalobos, H., Polishuk, A., and Gamer, S. (2022) "Augmented and virtual reality experiences for learning robotics and training integrative thinking skills", Robotics, 11(5), Article 90.

Yang, C., Zhang, J., Hu, Y., Chang, M., and Wang, S. (2024) "The impact of virtual reality on practical skills for students in science and engineering education: A meta-analysis", International Journal of STEM Education, 11, Article 28.

Yu, Z. (2022) "A meta-analysis and systematic review of the effect of virtual reality technology on users' learning outcomes", Computer Applications in Engineering Education, 30(3), pp. 884–896.