VIVARIA: a Desktop Hand-Tracked Virtual Laboratory for 3Rs-Aligned Animal Procedure Training

Massimo Verzeni[1], **Federica Riva**[2], **Benedetta Grossi**[3], **Francesco Migliavacca**[3], **Luca Mainardi**[1], **and Pietro Cerveri**[1,4]

[1] Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

[2] Department of Veterinary Medicine and Animal Science, Università degli Studi di Milano, Milan, Italy

[3] Department of Chemistry, Material and Chemical Engineering, Politecnico di Milano, Milan, Italy

4 Department of Industrial and Information Engineering, Università di Pavia, Pave, Italy

E-mail: `pietro.cerveri@unipv.it`

Abstract.

Purpose: This study investigated whether a desktop-based virtual environment using markerless hand tracking can effectively support the training of laboratory animal facility procedures. The objective was to assess system usability and to identify precision-critical phases of task execution through detailed interaction analysis.

Methods: The simulator, developed in Unity, utilized Ultraleap markerless hand-tracking sensor. The system implemented a modular task/step/condition architecture and automatically logged objective interaction metrics, including completion time, errors, rewinds, and grasp attempts. Twenty-six participants performed three simulated procedures (mouse cage transfer, jaw blood collection, and intraperitoneal injection). Performance logs were analyzed together with post-session questionnaire responses assessing usability and perceived educational value.

Results: Completion times differed substantially across procedures. The mouse cage transfer task was completed rapidly (27.99±15.05 s), whereas the jaw blood collection task required significantly longer execution times (124.72±76.09 s) than the intraperitoneal injection task (86.16±40.35 s; $p = 0.017$). Critical errors were concentrated in the precision-demanding procedures, with 19 errors recorded in the jaw blood collection task and 10 in the intraperitoneal injection task, while no errors occurred during cage transfer. Step-level analysis further highlighted the difficulty of the jaw blood sampling phase, which showed a success rate of 46.15%, compared with 76.92% for the injection step and near-perfect success for preparatory steps.

Conclusion: The results suggest that desktop-based virtual laboratories using markerless hand tracking can provide a viable preparatory environment for laboratory animal science training. While precision-sensitive interactions remain challenging, the system successfully reproduced the relative complexity of the simulated procedures and enabled detailed analysis of interaction behavior. Such platforms may contribute to training approaches aligned with the principles of Replacement, Reduction, and Refinement (3Rs) by allowing users to practice procedural workflows before interacting with live animals.

.

Keywords: Virtual Reality, 3Rs, laboratory animal science, interactive 3D exploration.

1. Introduction

Education and training in laboratory animal science (LAS) have increasingly been shaped by ethical, scientific, and regulatory requirements emphasizing competence prior to performing procedures on live animals [1]. The 3Rs framework (Replacement, Reduction, Refinement) has evolved from an ethical guideline into an operational principle embedded within governance structures and educational standards [2]. Within the European Union, Directive 2010/63/EU promotes refinement of procedures and the development of alternative approaches, positioning structured training and competence assessment as central components of responsible animal use [3]. In this context, simulation-based education has gained relevance as a preparatory stage that can reduce avoidable animal use during early learning phases while improving procedural readiness [4, 5]. LAS education literature explicitly recognizes the role of digital and immersive technologies in supporting 3Rs-aligned training pathways [1, 6], while advances in computational modeling and "virtual animals" further reinforce the role of digital complements to traditional experimentation [7, 8]. Across healthcare and veterinary education, immersive Virtual Reality (VR) has demonstrated positive effects on knowledge acquisition and skill development [9, 10, 11]. However, systematic reviews report heterogeneity in study design, limited standardization of pass–fail criteria, and insufficient evidence of reliable skill transfer [12]. These findings indicate that, before claims of pedagogical efficacy can be made, foundational aspects such as interaction reliability and usability require rigorous validation.

In this context, interaction modality is a key factor determining whether virtual training can support precision-critical workflows. Controller-based VR systems provide stable input and often achieve high performance in selection tasks [13], but introduce abstraction in procedures requiring fine motor control and anatomical realism [14]. Markerless hand tracking enables direct mapping between real and virtual hand motion, increasing ecological validity and immersion [15], but remains affected by occlusion, tracking instability, and reduced fine-motor precision [16]. Comparative studies highlight these trade-offs, showing that no single interaction modality consistently outperforms others in precision-demanding tasks [17]. Two critical aspects remain insufficiently characterized in LAS procedural simulation. First, most interaction studies focus on immersive head-mounted display environments rather than desktop-based, markerless hand-tracked configurations offering scalable deployment [18]. Second, existing evaluations typically report aggregate task-level outcomes, without structured step-level performance logging capable of localizing precision-critical interaction bottlenecks [19]. As a result, while interaction trade-offs are well documented in VR research, the robustness of desktop markerless hand tracking for structured LAS workflows remains underexplored.

This work investigates whether a desktop-based markerless hand-tracked environment can reliably support precision-sensitive LAS procedures while enabling structured step-level performance logging. To address this question, we present the design and foundational usability evaluation of a desktop virtual laboratory aligned with the 3Rs framework. The platform, VIVARIA, simulates representative mouse-handling procedures, including manipulation, blood collection, and injection, within a virtual animal facility. Each procedure is structured as a step-wise task/step/condition workflow, enabling monitoring of interaction quality, classification of critical and non-critical errors, controlled rewind, and detailed step-level logging. This design allows systematic identification of precision-critical interaction bottlenecks under realistic constraints. Rather than claiming pedagogical effectiveness, this work focuses on validating interaction robustness and identifying precision-critical limitations that must be addressed before competency-based validation. Thus, the main contributions of this work are:

- a modular task/step/condition architecture with explicit error modeling and step-level logging;
- a stability-oriented desktop hand-tracking design supporting precision-sensitive interaction
- a step-level usability validation framework combining objective metrics and subjective data to identify interaction bottlenecks.

1.1. Related work

Early studies highlighted the potential of immersive and collaborative virtual environments to support learning and structured assessment in laboratory contexts [4]. More recently, LAS-specific VR initiatives demonstrated feasibility and high learner acceptance [20]. For example, immersive 360° VR modules were proposed to prepare users for in vivo procedures while supporting refinement and reduction in training [21], and collaborative VR platforms have been developed to enable multi-user interaction and pedagogical modeling in animal experimentation training [22]. These systems confirmed the relevance of VR for LAS training, but primarily focused on procedural familiarization, immersive exposure, or collaborative

Table 1. Comparative summary of hand interaction modalities for procedural simulation.

Modality	Strengths for training	Typical flaws	Representative evidence
Controller-based	Reliable discrete input; good performance for selection/manipulation in common VR UI styles	Abstract interaction (reduced ecological validity for finger-level dexterity); limited anatomical realism of hand posture	Controllers outperform free-hand for raycast tasks in speed/accuracy and exertion [14]
Markerless	Natural mapping of hand motion; potentially higher ecological validity for manual procedures	Tracking loss/occlusion; gesture ambiguity; variable robustness; fatigue risk in mid-air tasks	Tracking robustness and evaluation challenges [15]; mixed preferences depending on interaction style [17]
Haptics	Adds touch/force cues; improves realism; can improve performance in procedural training	High cost and integration complexity; hardware constraints; calibration and comfort issues	Haptic VR can improve surgical training outcomes [26]; haptics viable in complex operational training [27]

learning, rather than detailed modeling of interaction performance. Beyond LAS, simulation technologies have been widely adopted in veterinary and biomedical education. Reviews indicate that simulators can improve procedural competence while reducing reliance on live animals and cadaver material, although coverage across procedures remains uneven and evaluation practices heterogeneous [23]. Applied VR systems, including orthopedic training modules [24] and interactive simulation environments with embedded feedback mechanisms [5], demonstrated feasibility and positive learner reception. However, these studies also reported practical constraints such as technical instability, cognitive load, and cybersickness, and generally relied on task-level outcomes rather than fine-grained performance metrics. Systematic reviews further emphasized the lack of standardized competency measures and validated pass–fail criteria, as well as limited evidence of skill transfer to real-world contexts [12]. Interaction modality has been extensively investigated, highlighting trade-offs between controller-based, markerless, and haptic interaction in terms of precision, realism, and robustness (Table 1). Controller-based systems provide reliable input but introduce abstraction in fine motor tasks, while markerless hand tracking offers more natural interaction at the cost of reduced precision and stability [13, 15, 25, 17, 14]. Haptic systems improve realism and performance in precision tasks but require specialized hardware and increase system complexity [26, 27]. Overall, existing work demonstrates the educational potential of VR-based training but lacks structured approaches to analyzing interaction at the procedural level. In particular, current systems do not provide step-level performance logging capable of identifying precision-critical phases within complex workflows, motivating the approach proposed in this work.

Table 2. VIVARIA framework. Procedure, interaction types, namely grasping (G), repositioning (R), precision manipulation (P), error-state model as critical (C) vs. non-critical (NC), task description.

Procedure	Interaction	Error	Description
Login	—	—	User profile registration/retrieval
Practice	G/R	NC	Anatomy exploration and tool handling
Cage transfer	G	C	Transfer of mouse across cages
Ear biopsy	G/R/P	C	Tissue sampling with sample storage
Blood collection	G/R/P	C	Vascular puncture and sample collection
Injection	G/R/P	C	Abdominal needle positioning and syringe actuation
Euthanasia	G/R	NC	Suppression of the mouse by CO_2 gas

2. Methods

2.1. System Overview

The proposed system, called VIVARIA, modeled a virtual animal facility environment enabling the execution of representative mouse laboratory procedures. It was implemented as a desktop-based application in Unity (version 6000.2.10f1), using the universal render pipeline (URP) for real-time rendering. Interaction and manipulation were enabled through markerless hand tracking via Ultraleap 3Di‡ stereo infrared sensor (Ultraleap Limited, UK) operating at 90 Hz. A detailed digital mouse surface model, including male and female variants, was obtained from the commercial asset repository TurboSquid (model ID: 843255) and used as the anatomical basis for the simulated procedures. It included ten anatomical systems. VIVARIA integrated two main functional contexts: (i) an exploratory learning context devoted to practice three-dimensional anatomical visualization and object handling, and (ii) a laboratory context for the execution of the laboratory procedures. A main login/registration scene enabled recording basic amnestic subject data (Table 2). The exploratory learning context was devoted to acquire familiarity with a gesture vocabulary (single-hand and dual-hand operations), practicing interaction with and manipulation of the mouse and tools (needle, capillary tube, syringe, biopsy forceps, mouse cage). The procedural context envisaged five procedures, namely a) mouse cage transfer, b) ear biopsy, c) blood collection in three different sites (jaw, abdomen, tail), d) intraperitoneal injection, e) CO_2-based euthanasia. The selected procedures were designed to span a progressive range of interaction demands relevant to early-stage laboratory animal science training. In particular, mouse cage transfer represented a basic handling task, injection required

‡ Technical specifications available at`https://www.ultraleap.com/products/`

stable positioning and controlled tool placement, whereas blood collection involved fine anatomical targeting and sample acquisition. This progression was intended to reflect increasing levels of precision and motor control required in real laboratory procedures, thereby enabling systematic assessment of interaction difficulty [20]. Accordingly, each procedure was categorized using a combinations of three types of object interaction (grasping, repositioning, precision manipulation), along with an error-state model. Automatic calibration parameters defined the geometric relationship between the display, user, and sensor (screen height, tilt, and camera–screen distance), ensuring spatial coherence between physical hand motion and the virtual workspace. Subject-specific calibration was applied to scale the virtual hand–arm model and adjust camera distance and viewing angle, normalizing interaction conditions across users with different body sizes.

2.2. Software Architecture

The simulator adopts a modular, event-driven task/step/condition architecture designed to formalize procedural workflows and ensure scalability, reproducibility, and structured performance logging. Four main modules, Procedure, Interaction Handler, Runtime Orchestration, and Feedback, interact to drive the overall process (Fig. 1). Each procedure is defined as a *Task* composed of an ordered sequence of steps (Procedure Module). A task includes metadata, progression rules, and global performance tracking. Each *Step* encapsulates procedural instructions, optional reference video guidance, and a set of logical conditions governing state transitions. These *StepCondition* objects are implemented as reusable scripts that perform logic checks (e.g., tool placement, anatomical targeting, puncture location, and sequence compliance). Each condition returns one of three states: *in progress*, *completed*, or *failed*. Failures are explicitly classified as critical (C) or non-critical (NC). Within the Runtime Orchestration module, the *StepController*, through the *TaskManager*, handles condition evaluation. Critical failures trigger an automatic rewind mechanism managed by the *StateManager*, which restores object transforms and procedural states before execution resumes. Non-critical failures instead notify the *UIController*, which displays warning messages through the *UIWarnings* component without resetting the simulation state. User actions, detected via Ultraleap, are processed by the *ObjectHandler* within the Interaction Handler module. The handler translates physics- and trigger-based events into updates of the *StepCondition*, converting low-level interaction signals into semantically meaningful procedural states. The *StepController* continuously evaluates these states and coordinates task progression through the *TaskManager*. In parallel, the *GrabManager* monitors grasp interactions with the mouse model and tools, while the *ObjectHandler* triggers corresponding audio and visual feedback.

Given a selected task, the corresponding Unity scene is initialized by placing the relevant

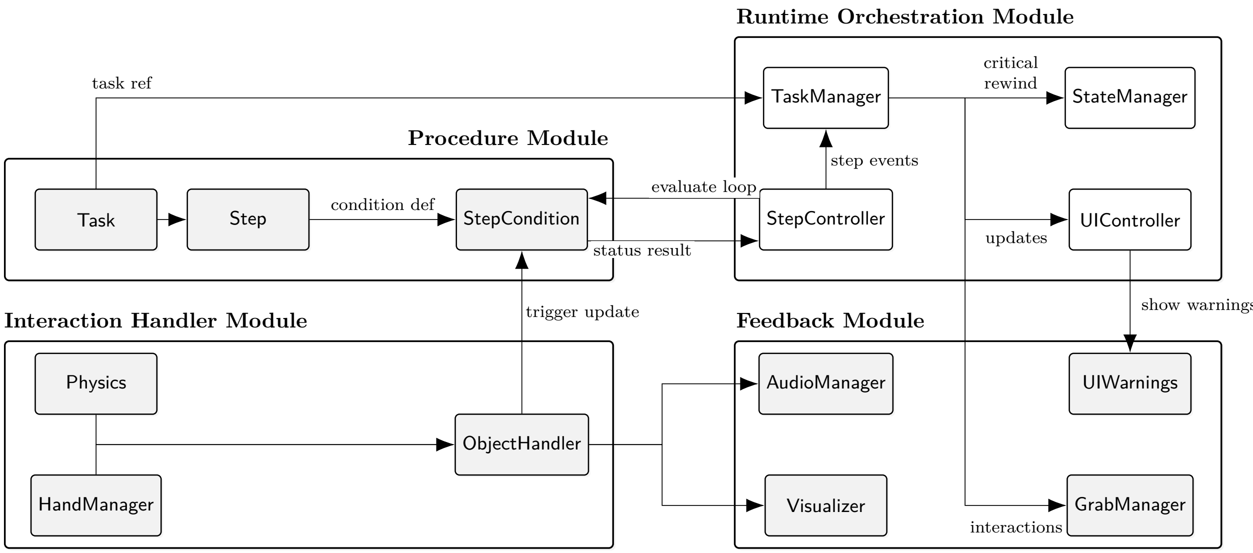

Figure 1. Component architecture of VIVARIA. The system decouples the procedure module from runtime orchestration, low-level interaction handlers, and user feedback.

graphical objects and interface elements, and by setting the camera view and orientation. Execution begins with user interaction detection via the *ObjectHandler* (Fig. 2). The handler updates the *StepCondition*. The *StepController* evaluates the condition state and forwards it to the *TaskManager*. If the step is successfully completed, the system updates the visual feedback and progresses to the next step. In case of a critical failure, the step is reset and the scene state is restored. For non-critical failures, a warning message is displayed without interrupting execution. This loop continues until the task is completed or interrupted, after which all performance metrics are stored in a JSON file.

2.3. Interaction Design and Stability-Oriented Strategies

A minimal gesture vocabulary was implemented, including pinch for fine manipulation, grab for tool handling, line pointing for anatomical identification, palm orientation for graphical user interface interaction, and wrist-touch for contextual menu activation. Given the known limitations of markerless hand tracking, several stability-oriented design strategies were adopted. Core instruments were configured as gravity-free kinematic rigid bodies to prevent unintended physics artifacts, while collision filtering restricted interactions to anatomically relevant targets through dedicated layers and trigger-based logic. Hand–object interference was reduced using Ultraleap's *IgnorePhysicalHands* module to suppress unintended collisions caused by dense finger colliders. To avoid unreachable states, released tools were automatically returned to predefined positions. Additional precision scaffolding included

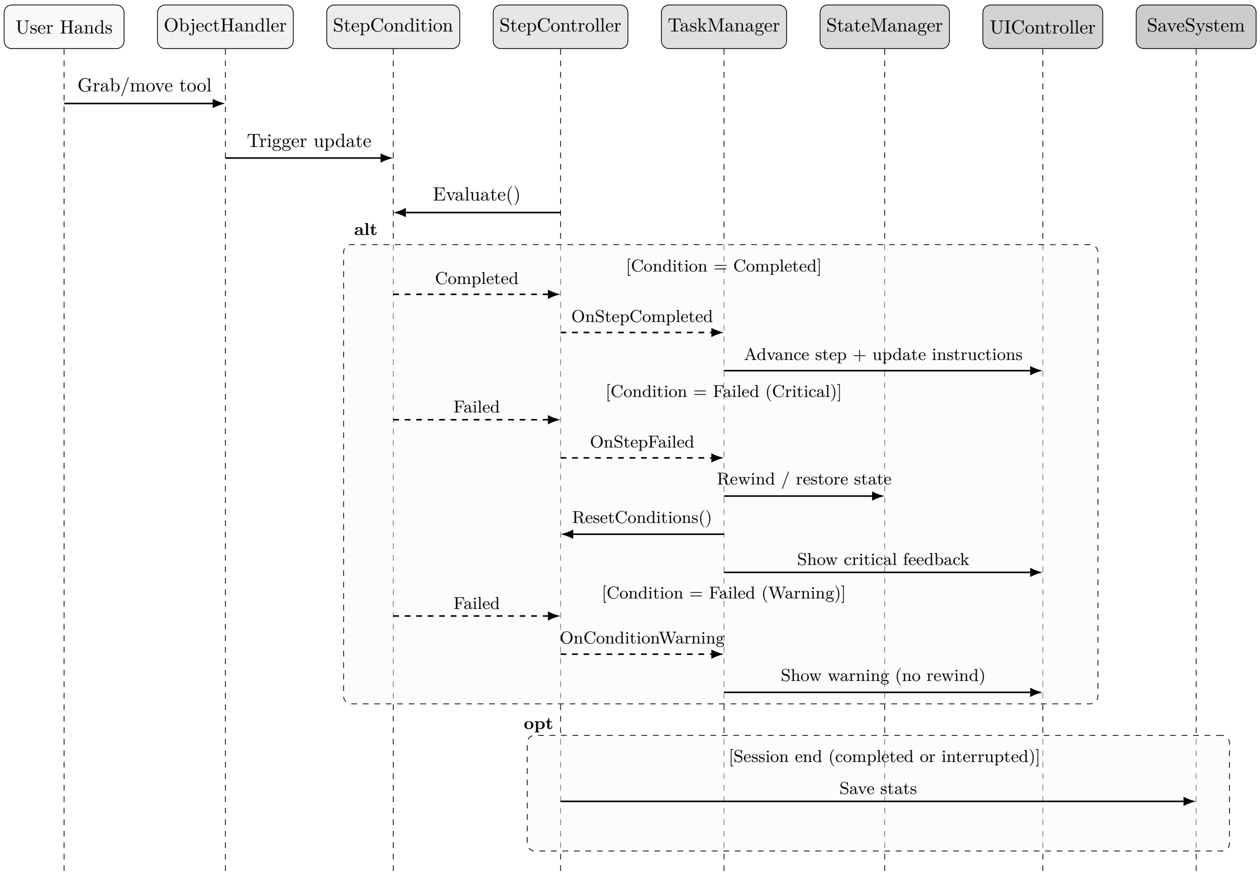

Figure 2. Event-driven sequence diagram of task execution. The StepController evaluates the state of StepCondition objects and triggers state progression, critical rewind via StateManager, or non-critical warning feedback via TaskUIController.

visual aids such as needle-tip projection beams and localized spherical transparency to support depth perception during vascular targeting. These strategies deliberately prioritized interaction stability over full rigid-body realism during the foundational validation phase.

2.4. Study Protocol

The study involved 26 students (aged 24.0±3.2 years, range 22–32), with a majority of females (76.9%), enrolled in the MSc program in Veterinary and Biotechnology Medicine (University of Milan, Milan, Italy). The study was conducted between late January and early March 2026 in accordance with the guidelines of the Politecnico di Milano Ethics Committee (Opinion 3/2019, 19/02/2019). All participants provided informed consent in compliance with GDPR regulations. Participants were seated approximately 30 cm from the device and were instructed to keep their hands within the sensor field of view, minimizing hand overlap. The Ultraleap device was connected to a desktop computer in an upward-facing configuration positioned directly beneath the monitor, following the manufacturer's recommended setup. The experimental protocol consisted of three sequential phases. First, a guided familiarization phase was conducted within the exploratory context, structured into two complementary submodules. The first submodule focused on gesture training, allowing participants to practice core hand and finger interactions, including zooming and rotating the mouse model, opening and navigating the interface menu, accessing anatomical visualization functions, performing clipping operations to section the model, and using pointing gestures to identify anatomical regions and display labels. The second submodule focused on object interaction, enabling participants to practice grasping and manipulating the virtual mouse and procedural instruments within a simplified environment. Second, participants performed three representative procedural tasks selected from the available set (cfr. Table 2): (i) mouse cage exchange (grasping/manipulation), (ii) jaw blood collection (precision targeting and sampling), and (iii) intraperitoneal injection (positioning and delivery). The tools introduced during the familiarization phase were used during task execution, including the needle for jaw puncture, the capillary tube for blood sampling, and the syringe for injection. At the beginning of this phase, the experimenter briefly described the in-task interface (video guidance, retry, and exit functions). No strict time limit was imposed; however, participants were allowed to proceed to the next task if execution exceeded 5 minutes. Third, immediately after completing the tasks, participants provided subjective feedback through a structured questionnaire.

2.5. Interaction metrics

During the second test session, objective performance metrics were automatically logged through a profile-based persistence module without experimenter intervention. At the task level, completion time t_T was defined as the elapsed duration between task initiation and either successful completion or interruption. The number of task attempts was computed as $A_T = 1 + N_{\text{rewind},T}$, where $N_{\text{rewind},T}$ denotes the number of state resets triggered during execution. At the step level, attempt counts were defined as $A_{S_i} = 1+N_{\text{rewind},S_i}$, and step-level success rates were computed as the proportion of step executions completed without critical errors or manual rewinds. Confidence intervals (95%) were estimated assuming a binomial model. Interaction logs also captured object manipulation events. For each task, the number of successful (G_{success}), failed (G_{failed}), and total grasp attempts ($G_{\text{total}} = G_{\text{success}} + G_{\text{failed}}$) were computed, providing an indirect measure of interaction difficulty. Descriptive statistics were reported as mean $\pm$ standard deviation (SD). Task-level comparisons were performed using paired statistical tests. Effect size was quantified using Cohen's d for paired comparisons between tasks. Associations between subjective questionnaire responses and objective metrics (e.g., completion time, grasp attempts, errors, and rewinds) were evaluated using Spearman correlation coefficients, with significance set at $p < 0.05$.

2.6. Questionnaire-based analysis

The questionnaire was based on a five-level Likert scale. All items were formulated in a positive direction to avoid response bias associated with reverse-coded statements. Questionnaire items were formulated to align directly with the system's interaction constraints (e.g., precision handling, gesture recognition stability) and were analyzed descriptively. Overall, the questionnaire comprised 17 questions distributed in four domains: (i) background experience to asses prior familiarity with interactive technologies and virtual reality (2 questions), (ii) perceived learning outcomes as a self-reported understanding of the procedural content (2 questions), (iii) usability and interaction experience to assess gesture intuitiveness, object manipulation control, task progression easiness, and feedback clarity (9 questions), and (iv) realism and educational relevance to asses the perceived degree of realism of the environment, and educational value within a laboratory animal science training context (4 questions). As the primary aim of the study was exploratory usability validation rather than psychometric scale development, no formal factor analysis was conducted. Questionnaire responses were analyzed descriptively and, where appropriate, correlated with objective performance metrics (e.g., task completion time and step-level attempts) to support a mixed-method evaluation framework. To investigate the relationship between subjective perception and objective performance, correlations between questionnaire responses and interaction metrics were computed at the participant level. Likert-scale responses were converted to numerical values (1–5), and non-parametric Spearman correlation coefficients were used to evaluate associations between subjective ratings and objective performance measures (e.g., completion time, grasp attempts, and critical errors). Statistical significance was assessed at $p < 0.05$.

3. Results

3.1. VIVARIA scenes

Representative VIVARIA exploratory scenes are illustrated in Figs. 3, 4. As earlier described, the practice module enabled participants to engage with core hand gestures, for instance double-hand pinch for mouse model zooming, and single-hand pointing to anatomical structure labeling (Fig. 3). Likewise, participants practiced grasping and manipulation of the virtual mouse and procedural instruments (Fig. 4). Representative VIVARIA procedure scenes are illustrated in Figs. 5, 6. The jaw blood collection task was characterized by three main sequential interaction steps, including needle alignment, puncture, and sample collection (Fig. 5). Intraperitoneal injection task required the alignment, insertion of the syringe into the abdominal region, and plunger actuation simulating the compound delivery (Fig.6). Across scenes, visual and audio feedback were provided to guide interaction and indicate action completion.

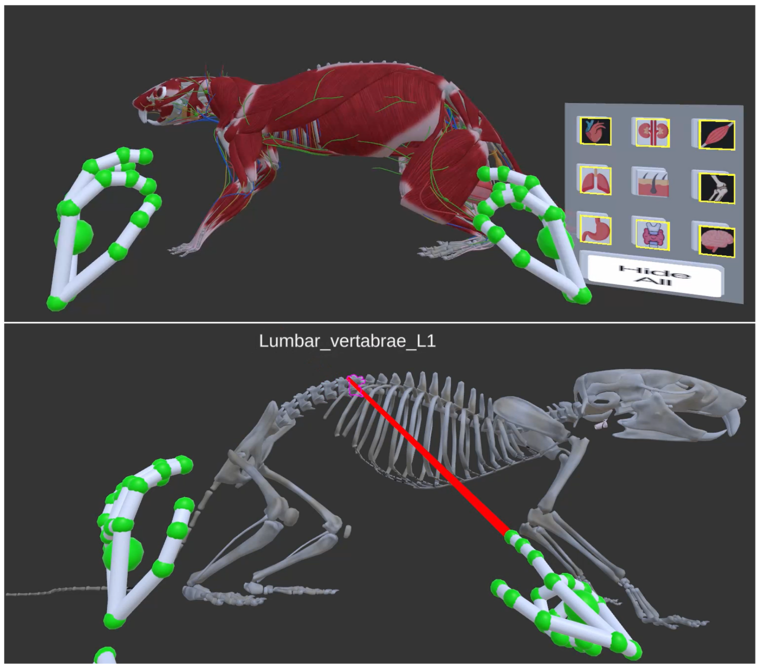

Figure 3. Two examples of virtual environment practice. (top) The user may manipulate the mouse (zoom, rotation, translation) and visualize different anatomical systems. (bottom) The user can interact with the mouse model to outlines specific anatomical part/regions by index finger pointing.

Table 3. Task-level performance metrics across the three simulated procedures.

Task	Time (s)	Critical errors	Rewinds
Mouse cage transfer	27.99±15.05	0.00±0.00	0.00±0.00
Jaw blood collection	124.72±76.09	0.73±1.15	0.19±0.49
Intraperitoneal injection	86.16±40.35	0.38±0.80	0.04±0.20

Table 4. Object grasp interaction statistics across the three simulated procedures.

Task	Mean success	SD	Mean failed	SD	Mean total	SD
Mouse cage transfer	3.85	2.73	0.92	1.41	4.77	3.83
Jaw blood collection	16.00	9.35	8.38	8.49	24.38	14.36
Intraperitoneal injection	15.62	7.50	3.31	4.22	18.92	10.03

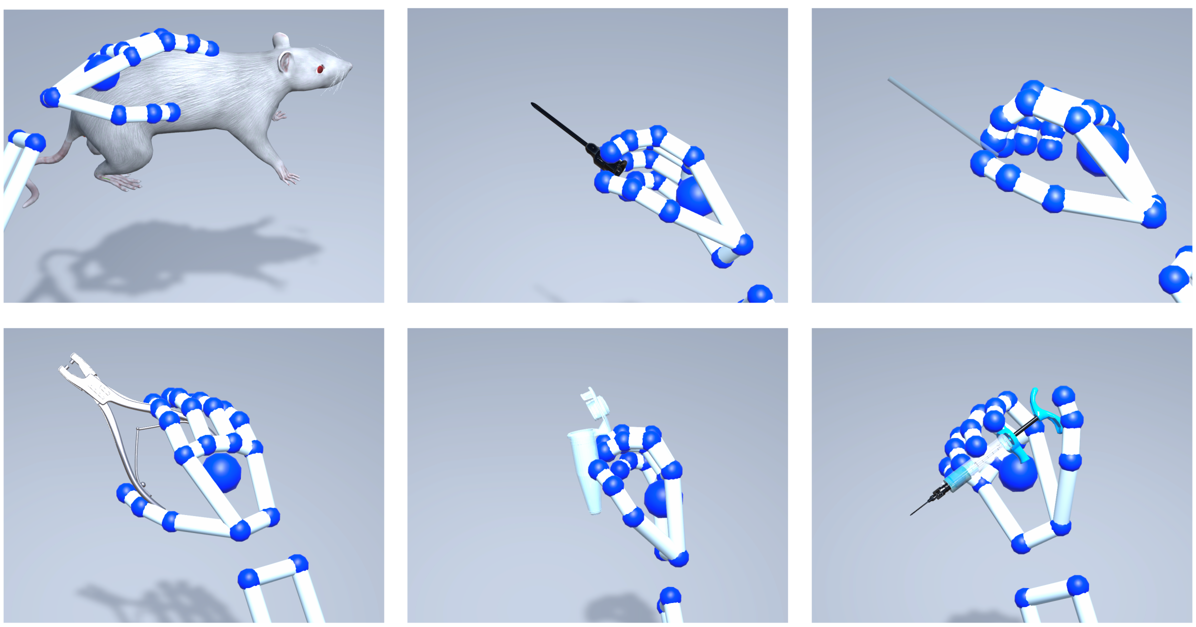

Figure 4. Mouse and main procedural instruments as needle, capillary tube, biopsy punch, Eppendorf tube, and syringe.

3.2. Quantitative performance

3.2.1. Completion time and procedural errors: Completion times differed substantially across the three simulated procedures (Table 3). The mouse cage transfer task was completed rapidly by most participants (27.99±15.05 s), indicating that this procedure was relatively straightforward within the virtual environment. In contrast, the jaw blood collection task required considerably longer execution times (124.7±76.09 s), reflecting higher procedural complexity and greater variability in participant performance. The intraperitoneal injection task showed intermediate completion times (86.16±40.35 s), suggesting a moderate level of task difficulty. Completion times also differed significantly between the two precision-demanding procedures. The jaw blood collection task required significantly longer execution times than the intraperitoneal injection task ($p = 0.017$). This difference was consistent with the increased interaction difficulty associated with targeting a small anatomical region in the mouse jaw compared with the larger abdominal region required for intraperitoneal injection.

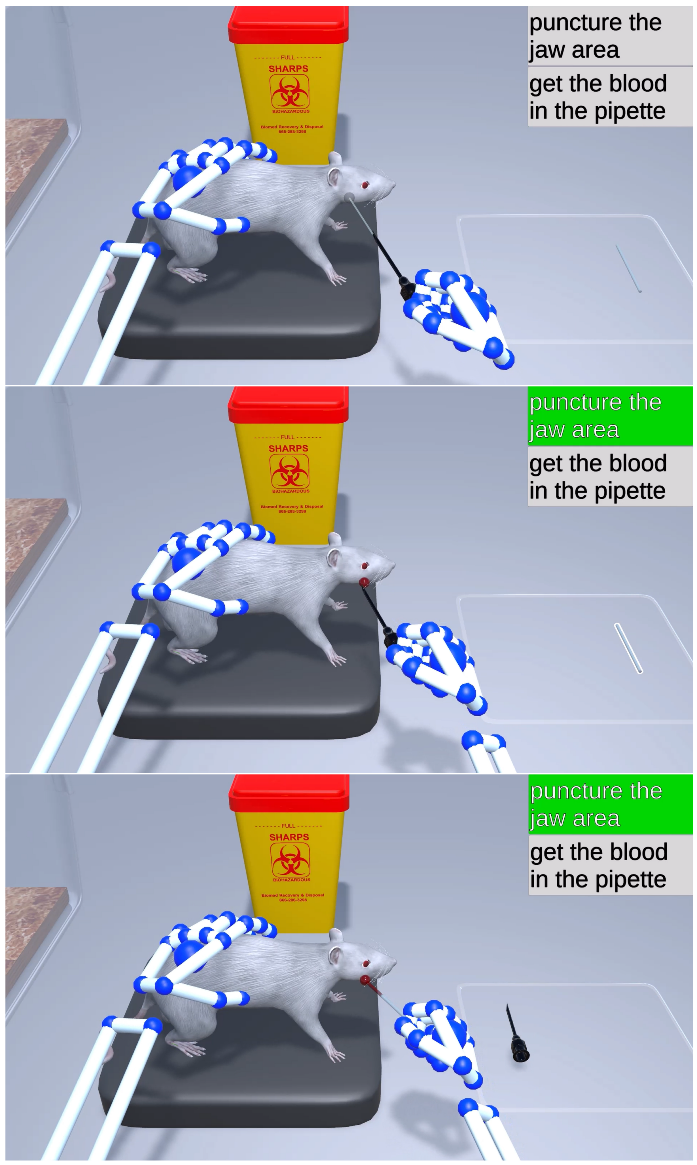

Figure 5. Jaw blood sampling procedure. (top) Needle alignemtnwith the jaw. (middle) Once the puncture was performed a blood drop appeared. (bottom) Capillary tube sampling.

No critical errors or rewinds were observed during the mouse cage transfer task, indicating that participants were able to complete this procedure reliably without major interaction difficulties (Table 3). In contrast, the jaw blood collection task produced the highest number of critical errors (0.73±1.15) and the largest number of manual rewinds (0.19±0.49). The intraperitoneal injection task showed intermediate values for both critical errors (0.38±0.80) and rewinds (0.04±0.20). Overall, a total of 19 and 10 critical errors were observed during the jaw blood collection and intraperitoneal injection task, respectively. These results were consistent with the completion time analysis, as the procedures requiring greater interaction precision also exhibited longer execution times and a higher frequency of errors.

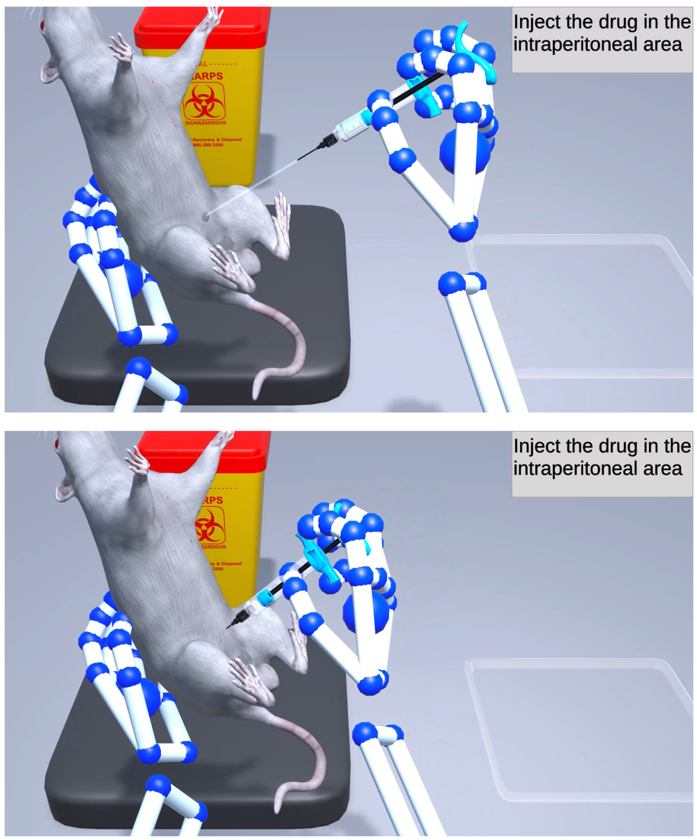

Figure 6. Intraperitoneal injection procedure. (top) Alignment of the syringe with the mouse abdomen; a visual guide (light pencil) assists the user in correctly orienting the syringe tip. (bottom) The user inserts the needle into the abdominal region, activating compound delivery.

3.2.2. Step-level analysis: Step-level analysis further highlighted differences in procedural difficulty across specific interaction phases. Most preparatory actions, such as removing the mouse from the cage and placing it on the workbench were completed successfully by nearly all participants, with success rates close to 100%. In contrast, grabbing the three small tools (needle, capillary needle and syringe) appeared to introduce additional interaction difficulty. As expected, the jaw blood sampling step showed the lowest success rate (46.15%, 95% CI [0.29, 0.65]), indicating that this phase represented the main source of interaction difficulty. This step required participants to hold and orient the mouse so that the head orientation was consistent for accurate needle targeting. The intraperitoneal injection step showed a higher but still imperfect success rate (76.92%, 95% CI [0.58, 0.89]), reflecting the need to stabilize the animal while exposing the abdominal region to allow proper syringe insertion. In contrast, the mouse cage transfer task showed perfect success rates for both steps (100%), confirming that this procedure required minimal precision-based interaction.

The quantification of the number of successful and failed grasping actions further revealed differences in object manipulation behaviour (Table 4). The mouse cage transfer task required only a limited number of grasping actions, with participants interacting mainly with the mouse (3.62±2.47 times) and the cage (1.50±0.71 times). In contrast, the jaw blood collection task involved substantially more object manipulations. Participants interacted frequently with the mouse (10.54±7.28 times) and the needle (3.85±2.94 times). This higher interaction frequency was largely associated with the need to properly hold and orient the head mouse to allow accurate targeting of the jaw with the needle. A similar interaction complexity was observed in the intraperitoneal injection task, where participants manipulated the mouse on average 11.08 times (SD = 7.45) and the syringe 3.46 times (SD = 1.94), with additional cage interactions (2.33±1.63 times). In this case, the main difficulty emerged from the need to hold the mouse in a stable position while exposing the abdomen sufficiently to enable correct needle insertion. Consistently with these procedural requirements, tasks involving precise positioning of the animal generated a larger number of grasp attempts (Table 4). The jaw blood collection task showcased the highest number of total attempts (24.38± 14.36 times), followed by the intraperitoneal injection task (18.92±10.03 times), whereas the cage transfer task required substantially fewer attempts (4.77±3.83 times). Effect size analysis revealed a large difference between the jaw blood collection and intraperitoneal injection tasks in completion time (Cohen's $d = 0.82$) and a moderate difference in grasp attempts (Cohen's $d = 0.42$), indicating substantially higher temporal and manipulation demands in the jaw blood collection procedure, likely due to the need for accurate targeting of a small anatomical region and stable orientation of the animal.

3.3. User Evaluation

3.3.1. Questionnaire results: Participants generally reported a positive evaluation of the virtual laboratory system across most questionnaire items (Fig. 7, Table 5). Several aspects

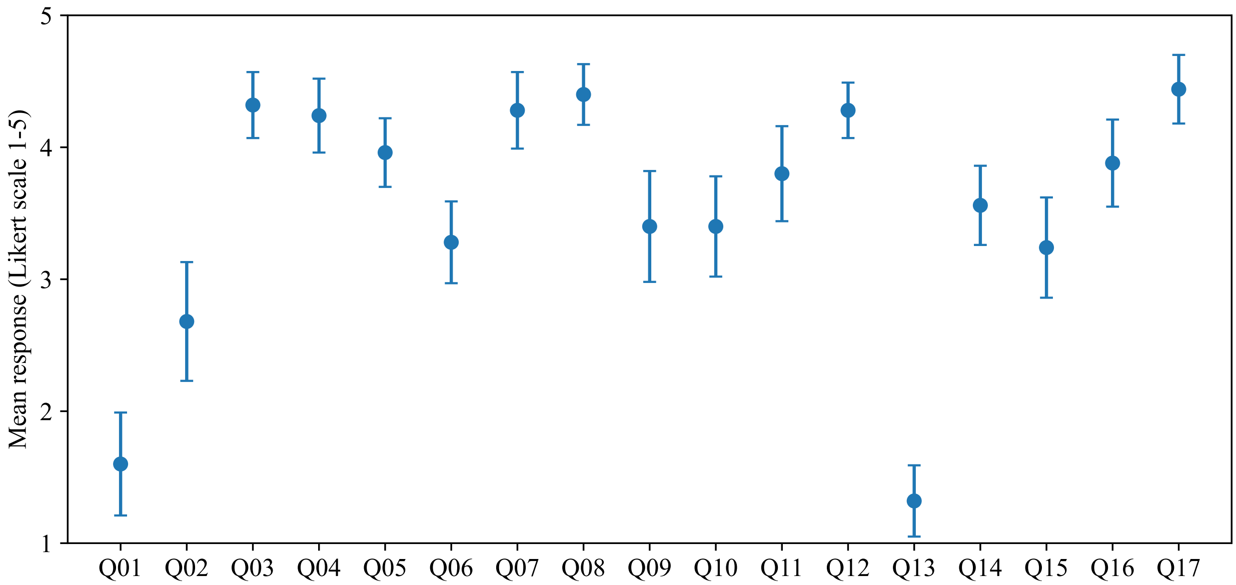

Figure 7. Mean participant responses for each questionnaire item (Q01–Q17) on a five-point Likert scale. Error bars denote 95% confidence intervals. The plot summarizes perceived usability, interaction quality, and educational value of the virtual laboratory system. Question identifiers (Q01–Q17) correspond to the items listed in Table 5.

related to usability and instructional value achieved high agreement levels. In particular, the clarity of visual and audio feedback provided by the system (Q08) received a mean score of 4.42 (SD = 0.58), with 95.83% of participants selecting *Agree* or *Strongly agree*. Similarly, the perceived usefulness of integrating the simulator into formal training programs (Q17) achieved the highest overall rating (mean = 4.46, SD = 0.66), with more than 91% of responses in the two highest Likert categories. Items related to procedural understanding and task progression (Q03–Q07) also received consistently high scores, with mean values ranging between 3.96 and 4.33, and agreement ratios exceeding 90% for several of these items. A comparatively lower score was observed for the item evaluating the smoothness and controllability of the hand-tracking interaction (Q06), which obtained a mean value of 3.25 (SD = 0.79) and an agreement ratio of 37.5%. This distribution indicated a more heterogeneous perception of the interaction stability among participants compared with other usability-related items. A comparatively lower score was observed for the item evaluating the smoothness and controllability of the hand-tracking interaction (Q06), which obtained a mean value of 3.25 (SD = 0.79) and an agreement ratio of 37.5%. This distribution indicated a more heterogeneous perception of the interaction stability among participants compared with other usability-related items. Participants reported very limited prior familiarity with

Table 5. Descriptive statistics for questionnaire items. Means are reported on a 1–5 scale. Agreement ratio represents the ratio between Agree and Strong agree responses.

Question	Mean	SD	95% CI	Agreement ratio
Q01	1.50	0.86	0.40	–
Q02	2.62	1.13	0.45	25.0%
Q03	4.33	0.64	0.25	91.67%
Q04	4.25	0.74	0.29	91.67%
Q05	3.96	0.69	0.28	83.33%
Q06	3.25	0.79	0.32	37.5%
Q07	4.29	0.75	0.30	91.67%
Q08	4.42	0.58	0.23	95.83%
Q09	3.38	1.10	0.44	50.0%
Q10	3.38	0.97	0.39	54.17%
Q11	3.79	0.93	0.37	70.83%
Q12	4.29	0.55	0.22	95.83%
Q13	1.33	0.70	0.28	–
Q14	3.54	0.78	0.31	62.5%
Q15	3.21	0.98	0.39	37.5%
Q16	3.88	0.85	0.34	66.67%
Q17	4.46	0.66	0.26	91.67%

immersive technologies (Q01: mean = 1.53, SD = 0.84), with only 5.26% of responses in the *Agree* or *Strongly agree* categories, indicating that most users approached the system without previous experience with virtual or interactive simulation environments. Despite this limited background, the system was generally well tolerated during the session. Reports of interaction-related discomfort were minimal (Q13: mean = 1.33, SD = 0.70), with no participants selecting the agreement categories for this item, suggesting that the desktop hand-tracking interaction paradigm did not induce noticeable physical symptoms such as fatigue or visual strain during the experimental session. Overall, the relatively narrow confidence intervals and the high proportion of agreement responses across multiple items indicated a consistent pattern of positive evaluations among participants.

3.3.2. Correlation analysis with quantitative results: Recall that questionnaire items were designed to capture distinct dimensions of the interaction experience. Specifically, Q06 assessed interaction controllability, Q09 user confidence during precision actions, Q10 perceived system stability, and Q08 feedback clarity. No significant correlations were observed between perceived interaction controllability (Q06) and objective performance metrics. In particular, Q06 showed no significant association with either the total number of grasp attempts ($\rho \approx -0.22$, $p = 0.47$) or total completion time ($\rho \approx -0.14$, $p = 0.65$) in precision-critical tasks. Similarly, confidence during precision-based actions (Q09) was not significantly correlated with the number of grasp attempts ($\rho \approx -0.24$, $p = 0.43$), indicating that self-reported confidence does not reliably reflect objective interaction performance. Perceived system stability (Q10) was also not significantly correlated with the number of critical errors ($\rho \approx 0.26$, $p = 0.39$), showing no clear relationship between subjective stability perception and error occurrence. In contrast, a moderate negative trend was observed between perceived feedback clarity (Q08) and the number of rewinds ($\rho \approx -0.37$, $p = 0.22$). Although not statistically significant, this trend suggests that clearer feedback may facilitate more efficient error recovery during task execution.

4. Discussion

4.1. Main findings

The present work focused on the design, development and usability evaluation of VIVARIA, a desktop-based virtual environment intended for the early-stage training of animal facility procedures. VIVARIA was conceived as a preparatory platform, to be integrated into a didactic curriculum, for students approaching these procedures for the first time, and its evaluation on participants with limited prior experience reflected this intended educational context. From a technical perspective, the adopted task/step/condition architecture represented a relevant contribution of the present work. The separation between low-level interaction handlers, procedural validation, and state management enabled controlled recovery after critical errors and supported fine-grained logging at the step level (see Figs. 1, 2), including completion times, errors, rewinds, and object manipulation events. The results indicated that the framework enabled procedural execution while exposing interaction limitations in precision-critical phases (see Fig. 7, Table 3). Basic handling actions were performed reliably, whereas tasks requiring stable animal positioning and accurate tool alignment remained more challenging. This confirmed that the system effectively reproduced the relative complexity of the simulated procedures, exposing the key interaction difficulties that characterize real-world laboratory tasks. Step-level analysis further clarified the origin of these differences. Preparatory phases were completed successfully by most participants, whereas precision-demanding phases exhibited lower success rates (see Fig. 4). The jaw blood sampling step highlighted the difficulty of stabilizing and orienting the mouse for accurate targeting, with a similar but less pronounced effect in the injection task. These precision-sensitive phases required larger completion time and more grasp attempts, showing higher variability in manipulation behaviour.

The relationship between subjective perception (see Fig. 7) and objective performance was further explored through correlation analysis. No significant associations were observed between perceived interaction controllability (Q06) and either grasp attempts or completion time, nor between perceived system stability (Q10) and the number of critical errors. Similarly, user confidence during precision-based actions (Q09) was not significantly associated with manipulation effort. Together, these findings indicate that subjective perception does not reliably capture objective interaction performance in precision-critical tasks. In contrast, a moderate negative trend was observed between perceived feedback clarity (Q08) and the number of rewinds, suggesting that clearer feedback may support more efficient error recovery. Although not statistically significant, this trend is consistent with the role of feedback in guiding user interaction. These results are consistent with the effect size analysis, which highlighted substantial differences in task difficulty despite the lack of strong associations with subjective perception. Overall, interaction difficulties appear to be primarily driven by task-specific and system-related constraints.

4.2. Comparison with the LAS literature

Previous LAS works have explored both computational approaches, including AI-based "virtual animals" aimed at reducing reliance on live-animal experimentation, and simulation-based environments for training in veterinary and laboratory animal science, primarily focused on procedural familiarization and learner engagement [20, 21, 5, 28, 8]. Tang et al. [20] demonstrated the educational value of immersive, gamified VR for laboratory animal handling training, particularly in supporting contextual learning, laboratory safety, and 3Rs-oriented reflection. However, their work primarily focused on pedagogical design and informal student feedback, without structured performance logging or step-level analysis. VR-based educational approaches, such as Lemos et al. [21], demonstrated the potential of immersive 360° environments to enhance learner preparation and reduce animal use, but remained limited to observational learning without direct interaction or quantitative performance assessment. Interactive VR systems with feedback loops have demonstrated their potential for training and assessment, primarily providing outcome-based evaluation [5]. However, they did not identify sources of interaction difficulty and analyze precision-critical manipulative bottlenecks. VR farm tours were shown to enhance student understanding of animal welfare and self-reported learning [28], they remained limited to observational experiences without quantitative performance assessment tough. By integrating procedural simulation with detailed interaction monitoring, VIVARIA enabled quantitative analysis of task execution and step level, identification of precision-critical interaction phases, and objective analysis of procedural difficulty.

4.3. Work Limitations

The absence of an immersive head-mounted display represents a first limitation of the current implementation. The use of a standard desktop configuration may reduce depth perception compared to fully immersive 3D visualization, potentially affecting spatial awareness and the ease of manipulation in precision-sensitive tasks. However, the desktop-based configuration was intentionally selected to maximize accessibility and facilitate integration into teaching curricula without requiring specialized hardware, thereby supporting scalability and ease of deployment. In addition, the system did not provide tactile feedback, which may limit realism in procedures requiring fine force control and precise instrument–tissue interaction. To compensate for this limitation, the environment incorporated multimodal guidance, including contextual textual cues, embedded instructional videos, and audio signals triggered by interaction events. These features supported user awareness and decision-making during task execution, partially offsetting the absence of haptic feedback, consistent with the high rating assigned to feedback clarity (Q08). Finally, the current implementation focused on a limited set of procedures representative of common laboratory workflows, which constrains the generalizability of the results to more complex scenarios encountered in a real animal facility. However, the three selected procedures were sufficient to capture a range of interaction conditions, from basic handling to precision-sensitive manipulation, and therefore provided an adequate testbed for evaluating interaction robustness and identifying procedural bottlenecks.

5. Conclusion

The present work focused on the design and usability evaluation of a desktop-based virtual laboratory environment intended for the early-stage training of laboratory animal science procedures. The system was conceived as a preparatory platform for students approaching these procedures for the first time, and its evaluation on participants with limited prior experience reflected this intended educational context. The proposed framework implemented a structured task/step/condition architecture that supported procedural guidance while enabling detailed logging of interaction metrics, including completion times, errors, rewinds, and object manipulation events. The desktop-based configuration was selected to maximize accessibility and facilitate integration into teaching curricula without requiring specialized hardware, thereby supporting scalability and ease of deployment. While the system effectively enabled users to gain familiarity with procedural workflows before entering the laboratory environment, precision-sensitive interactions remained challenging, particularly in phases requiring fine motor control and accurate tool positioning. In this sense, the proposed platform can support training approaches aligned with the principles of Replacement, Reduction, and Refinement (3Rs), by reducing the need for early-stage practice on live animals.

Acknowledgments and Declarations

- Funding: This research was supported by Piano Regionale Integrato della Sanità Pubblica Veterinaria 2019-2023 - Regione Lombardia, Italy
- Conflict of interest/Competing interests (check journal-specific guidelines for which heading to use): The authors declare no conflict of interest.
- Ethics approval and consent to participate: The study was conducted in full compliance with the principles of the Declaration of Helsinki and adhered to all applicable local regulations. It had the approval of Politecnico di Milano University ethics committee (Opinion 3/2019, 19/02/2019)
- Consent for publication: The authors affirm that participants provided informed consent for publication
- Data availability: The data presented in this study are openly available on https://github.com/Vivaria/ in StudyData folder
- Code availability: The data presented in this study are openly available on https://github.com/Vivaria/
- Author contribution: Conceptualization, F.R., M.V., and P.C. Data curation, B.G., M.V., Methodology, M.V., L.M., F.M., Supervision, F.R., and P.C. Writing-review and editing, F.R., F.M., L.M., and P.C.

References

[1] Nuno Henrique Franco, Angela Kerton, and David I. Lewis. Education in laboratory animal science and the 3rs. *Laboratory Animals*, 57(2):109–111, March 2023.

[2] Robert C. Hubrecht and Elizabeth Carter. The 3rs and humane experimental technique: Implementing change. *Animals*, 9(10):754, September 2019.

[3] I Anna S Olsson, Sandra Pinto da Silva, David Townend, and Peter Sandøe. Protecting animals and enabling research in the european union: An overview of development and implementation of directive 2010/63/eu. *ILAR Journal*, 57(3):347–357, May 2017.

[4] D. J. Ketelhut and S. M. Niemi. Emerging technologies in education and training: Applications for the laboratory animal science community. *ILAR Journal*, 48(2):163–169, January 2007.

[5] Lynn Keets, Pedro Boscan, Logan Arakaki, Benjamin Schraeder, Cyane Tornatzky, Marie Vans, Wenjing Jiang, and Sangeeta Rao. Interactive virtual reality with educational feedback loops to train and assess veterinary students on the use of anesthetic machine. *Journal of Veterinary Medical Education*, 51(3):412–420, June 2024.

[6] Lucy Whitfield. Watching me - watching you: Training and assessment in a virtual world. *Laboratory Animals*, 57(2):117–126, January 2023.

[7] Paloma Lobo Moraes, Lianna Ghisi, Anna Júlia B. Paes de Barros, Vithor Hugo de Carvalho Peixoto, Pedro Eduardo Brandini Népoli, Edson Moleta Colodel, Luiz Felipe Souza de Lima, and Roberto Lopes de Souza. Simuvet: a preliminary study of the innovative development of a simulator for epidural anesthesia training in dogs. *Frontiers in Veterinary Science*, 11, June 2024.

[8] Alireza Minagar and Mohammad Ali Sahraian. From rodents to algorithms: The rise of virtual animals in drug development and neuroscience. *Clinical Neurology and Neurosurgery*, 260:109234, January 2026.

[9] Feng-Qin Chen, Yu-Fei Leng, Jian-Feng Ge, Dan-Wen Wang, Cheng Li, Bin Chen, and Zhi-Ling Sun. Effectiveness of virtual reality in nursing education: Meta-analysis. *Journal of Medical Internet Research*, 22(9):e18290, September 2020.

[10] Hyeon-Young Kim and Eun-Young Kim. Effects of medical education program using virtual reality: A systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*, 20(5):3895, February 2023.

[11] Hinpetch Daungsupawong and Viroj Wiwanitkit. Virtual reality: Applications in veterinary science. *Veterinary Medicine and Science*, 12(2), February 2026.

[12] Karen Høgh Abrahamsen, Martine Siw Nielsen, Nicolaj Bruun Brandt, Anders Bo Nielsen, Lars Konge, Margit Dueholm, and Pernille Ravn. Training and assessment of competencies in hysteroscopy – a systematic review. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 318:114890, February 2026.

[13] Gavin Buckingham. Hand tracking for immersive virtual reality: Opportunities and challenges. *Frontiers in Virtual Reality*, 2, October 2021.

[14] Ricardo Pardo, Ignacio Miralles, Águeda Gómez-Cambronero, and Inmaculada Remolar. Analyzing the effectiveness and satisfaction of hand tracking vs. controllers among vr-experienced users. *Virtual Reality*, February 2026.

[15] Dennis Reimer, Iana Podkosova, Daniel Scherzer, and Hannes Kaufmann. Evaluation and improvement of hmd-based and rgb-based hand tracking solutions in vr. *Frontiers in Virtual Reality*, 4, July 2023.

[16] Anthony Steed and Jiaqi Lai. Comparison of hand tracking-based and controller-based interaction in a consumer virtual reality game. *Virtual Reality*, 29(3), July 2025.

[17] Tiffany Luong, Yi Fei Cheng, Max Möbus, Andreas Fender, and Christian Holz. Controllers or bare hands? a controlled evaluation of input techniques on interaction performance and exertion in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(11):4633–4643, November 2023.

[18] Nadia Savitri, Muhammad Wiranto Aris, and Ahmad Afif Supianto. Augmented reality application for science education on animal classification. In *2019 International Conference on Sustainable Information Engineering and Technology (SIET)*, pages 270–275. IEEE, September 2019.

[19] Emily L. Davidson, Kristina L. Penniston, and Walid A. Farhat. Advancements in surgical education: exploring animal and simulation models in fetal and neonatal surgery training. *Frontiers in Pediatrics*, 12, June 2024.

[20] Florence Mei Kuen Tang, Ray Mau Fung Lee, Roy Hok Lai Szeto, Justin Chak Ting Cheung, and Olivia Miu Yung Ngan. Experiential learning with virtual reality: animal handling training. *Innovation and Education*, 2(1):1–9, October 2020.

[21] Martin Lemos, Laura Bell, Susanne Deutsch, Leonie Zieglowski, Lisa Ernst, Daniel Fink, R. Tolba, Christian Bleilevens, and Julia Steitz. Virtual reality in biomedical education in the sense of the 3rs. *Laboratory Animals*, 57(2):160–169, October 2022.

[22] Mohamed-Amine Abrache and Lahcen Oubahssi. Towards a virtual reality collaborative platform for animal experimentation. In *2024 IEEE International Conference on Advanced Learning Technologies*

(ICALT), pages 232–236. IEEE, July 2024.

[23] Helen R. Braid. The use of simulators for teaching practical clinical skills to veterinary students — a review. *Alternatives to Laboratory Animals*, 50(3):184–194, May 2022.

[24] Katherine McCaw, Andrew West, Colleen Duncan, Danielle Frey, and Felix Duerr. Exploration of immersive virtual reality in teaching veterinary orthopedics. *Journal of Veterinary Medical Education*, 49(6):716–720, December 2022.

[25] Eric Godden, William Steedman, and Matthew K.X.J. Pan. Robotic characterization of markerless hand-tracking on meta quest pro and quest 3 virtual reality headsets. *IEEE Transactions on Visualization and Computer Graphics*, 31(5):3025–3034, May 2025.

[26] Abrar Gani, Oliver Pickering, Caroline Ellis, Omar Sabri, and Philip Pucher. Impact of haptic feedback on surgical training outcomes: A randomised controlled trial of haptic versus non-haptic immersive virtual reality training. *Annals of Medicine & Surgery*, 83, November 2022.

[27] Jared H. Hornsby, Austin Arnwine, Ashley Stowe, Mary Lin, Chris Mason, Mike Eakins, Devon Veller, Christian Mendoza, David Metcalf, Victor Oriaifo, Andrew Dunar, and Bob Crockett. Vr haptics for glovebox operations training. *Frontiers in Virtual Reality*, 6, March 2025.

[28] Nichole C. Anderson, Lauren Underwood, and Christopher J. Byrd. Student reported learning of swine and dairy welfare concepts following a virtual reality livestock farm experience. *Journal of Applied Animal Welfare Science*, pages 1–13, September 2025.