UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN
"SCIENZE VETERINARIE"

CICLO XXXVII

COMPARATIVE EVALUATION OF SIMULATOR-BASED AND CADAVERIC LIMB TRAINING FOR DISTAL LIMB NERVE BLOCKS IN THE HORSE: A STUDY OF OSCE PERFORMANCE AND STUDENT PERCEPTIONS IN TWO ITALIAN VETERINARY SCHOOLS

Coordinatore:
Chiar.mo Prof. Gaetano Donofrio

Tutori:
Chiar.ma Prof. Marina Martano
Chiar.mo Prof. Andrea Bertuglia

Dottorando: Dott. Marcello PALLANTE

Anni Accademici 2021/2022 – 2023/2024

SUMMARY

INTRODUCTION

Education in Veterinary Medicine was traditionally composed of frontal lectures and practical training on live animals or cadaveric parts. Nowadays, this approach is considered outdated, and Simulation-Based Medical Education (SBME) is gaining importance in the curricula of many Veterinary Universities around the world.

This new teaching method allows students to gain experience in a controlled, standardized, and risk-free environment. However, the motivations behind this trend are broader and multifaceted. First, SBME has been recognised to improve clinical skills by providing a safe and controlled environment where students can practice and refine their skills, with the possibility of repeating procedures until they gain the necessary confidence for professional practice. Moreover, this method enables to address ethical concerns regarding the use of live animals or cadavers. [1–3]

Specifically, there is an increased interest in reducing the use of cadavers and live animals for teaching purposes. Principles such as the 4Rs (replacement, reduction, refinement, and respect) and the concept of "never the first time on a live animal" are upheld as fundamental by many institutions, including the EAEVE (European Association of Establishments for Veterinary Education) and the AAVMC (American Association of Veterinary Medical Colleges). In 2021, the AAVMC established a task force of experts to develop the Guidelines on the Use of Animals in Veterinary Education. This task force, composed by experts from around the world, published a collection of guidelines in 2022 and an extensive applicative handbook for these guidelines in 2024, aimed at supporting and promoting humane and ethical animal use in veterinary education, guided by the principles of the 4Rs. In particular, the handbook suggests considering the use of simulators as a valid alternative to cadavers or live animals, albeit with certain limitations. [4–6]

Furthermore, it is important to consider the standardization of the training that simulators offer, ensuring each student receives the same learning opportunity, including simulations of complex or rare cases and procedures [1–3]. It is becoming

evident that the mere observation of medical or surgical procedures is not sufficient to stimulate problem-solving ability and critical-thinking, abilities that are indispensable for developing clinical skills [7].

On this topic, the EAEVE Strategic Plan 2020-2025 also emphasizes the importance of standardization in veterinary education through several strategic goals, such as driving harmonization of veterinary education, supporting the establishments in developing and strengthening their educational program, and monitoring quality standards including consistency. [4]

A 2004 paper published by B. Zemljic highlight that, in the current globalized economic and social scenario, Veterinarians should provide the same level of service and protection to every client, consumer and patient worldwide. Zemljic states that to meet this requirement, every veterinary student in the world should have the possibility to achieve the same minimum educational standards. To attain this goal, an independent control over the educational process, tailored to the needs and demands of users, is necessary. It is therefore necessary for teaching institutions to undergo a process of internal and external evaluation, conducted in a transparent, open, and independent manner, to ensure compliance with the required standards. [8]

Many studies have been conducted to investigate the effectiveness of simulators and their capacity to enhance clinical skills and confidence in medical students. Moreover, a wide range of simulators is available for human medical training, developed using the more up-to-date technological advancements, with numerous validation studies supporting their efficacy. Similarly, in veterinary medical education, some validation studies are currently available [1,9,10]. Of particular interest is the 2022 study by Noyes J. A. et al., which compared the effectiveness of simulator training versus traditional training in veterinary education. This meta-analysis indicated that simulator training can effectively enhance knowledge and clinical skills [1]. Another 2022 review by Helen R. Braid focused on the use of practical simulators for teaching clinical skills. This paper assessed the available simulators for teaching Day One Competences to veterinary students and discussed their efficacy. Braid’s study emphasized that teaching

with simulators can produce more confident graduates with greater proficiency in manual skills, protect animal welfare, and consequently reduce the need for cadavers and live animal in teaching [9].

Braid also underscored the role of the simulators in creating a safe environment where students can practice without the concerns for animal welfare or the stress of being observed by clients, both of which can inhibit learning and decision-making [9,11,12]. From Braid's literature review, a notable species discrepancy emerged, as only 12.3% of the published studies regarding animal simulators (excluding virtual simulators and communication- and scenario-based simulators) focused specifically on horses. These simulations included intravenous/intramuscular injections [13], intra-articular injections [14], nerve blocks [15], upper airways endoscopy [16], gynaecological examination [17,18], laparoscopic ovariectomy [19], and cardiac dissection [20].

With the introduction of SBME, not only do training methods gain an important new tool, but, on the other hand, assessment of competence should also be re-evaluated. This is driven by the increasing demand for more competent (veterinary) surgeons and elevated standards of the university institutions (such as the introduction of Day One Competences and quality assurance systems) and also by the needs of the students themselves.

Wass et al., in 2001, highlighted how medical students often feel overloaded and tend to focus on what is required for exams, sometimes overlooking the refinement of practical skills. According to Wass et al. (2001), "assessment drives learning. …assessment is the most appropriate engine on which to harness the curriculum" [21], and this applies equally to veterinary students.

The assessment of skills in the veterinary curriculum is critical for setting standards, driving learning, providing feedback to students, and preparing competent veterinary professionals. The use of the Objective Structured Clinical Examination (OSCE), adapted from human medicine, has proven effective in reliably assessing clinical skills and progression on the learning curve. This examination method involves standardized stations to assess different skills, ensuring consistency and reducing variability in

assessment. OSCEs have shown to enhance learning by linking the assessment stations to essential clinical skills. They can be used both formatively (to practice skills) and summatively (for final grading) [1,22].

In practice, the examination of students' skills using the OSCE method should ideally be organised into several stations, where students rotate according to a defined schedule. At each station, the candidate receives precise written instructions outlining what is expected of them. Each station should have a different, previously trained rater, who completes an objective-structured checklist for each candidate without providing any feedback during the examination. Feedback is provided only during a subsequent session.

From the students' perspective, OSCEs are generally perceived as valid and fair. Students recognize their relevance to real-world clinical practice, although they sometimes express a preference for live animal training for greater realism [23]. Given that simulators will eventually replace live animals teaching in the near future, more simulators, particularly those focused on large animals, and even more realistic ones, need to be developed and validated.

The purpose of this thesis is twofold: first, to investigate the practical and perceived educational outcome of a commercially available equine distal limb simulator to train the students in nerve blocks compared to traditional training on cadaveric legs; second, to begin the development of a proprietary 3D-printed high-fidelity model of the equine distal limb (up to the metacarpus) capable of simulating nerve blocks and other procedures, such as ultrasound examination on flexor metacarpal region, as part of the 2022 PRIN[1] "WELL-FARE: for animal welfare revolution in ethical education" project.

Focusing on this last topic, the project was conceived based on the consideration that Veterinary Schools are called to strengthen formal education towards the application of appropriate professional attitudes and ethical practices. This is requested also from

[1] PRIN 2022 - https://prin.mur.gov.it/ - Accessed 20/10/2024

the European Parliament action plan for transition to animal-free innovation [24], and the Italian Constitution that has recently changed to include animal welfare and protection as national obligations [25]. However, acquiring basic clinical competencies – also referred to as Day One Competences (DOCs) (e.g. performing an injection, taking a biopsy, performing a basic surgical intervention) – is performed on animals, cadavers, or body parts. Approximately 124 000 to 172 000 animals were used for educational and training purposes across Europe between 2014 and 2018 [26], with unaffordable costs for institutions and educational facilities and being perceived as ethically unsustainable [27,28]. Also, the European Parliament action plan for transition to animal-free innovation is recommending the transition toward an educational plan based on the so called "3R": Reduce the number of used animals up to zero, Refine the teaching setting without cadavers or part of animals, and Replace the traditional teaching methods with innovations that are realistic and tested for efficacy [29,30].

Among domestic species of veterinary interest, horses present the most significant potential benefits from the introduction of alternative teaching methods. In many cases, students in our country lack prior experience with horses and often feel uneasy when initially interacting with them, which hinders their ability to gain hands-on experience. Therefore, these new educational approaches should be carefully designed and supported within a structured learning environment to ensure their successful integration into the curriculum.

The choice of focusing on the development of a new simulator of the equine distal limb has been taken considering one of the most frequently encountered problems in equine practice: lameness. Being able to correctly identify the source of lameness is one of the fundamental skills for veterinary students [31].

MATERIALS AND METHODS

Study design

The student of the fourth year of the Veterinary Medicine course of the University of Parma and Turin were enrolled as the cohorts of study population. No exclusion criteria were applied, as the experiments were conducted during the curricular teaching sessions and no prerequisite exams were required to attend the practical sessions that were the focus of this study.

For the students from both cohorts, it was their first experience with the OSCE method of practical competency evaluation.

A total of 78 students participated in the study: 41 from the University of Parma and 37 from the University of Turin.

Two practical training sessions were held in the two Italian Veterinary Schools.

Prior to each practical training session, the students attended a 2-hour lecture on lameness examination, with special focus on nerve blocks. The lecture was delivered by an associate professor at the University of Turin, ACVSMR[2] diplomate.

For the practical training session, each class was divided into two groups: Group C consisted in students trained on cadaveric legs, and Group S consisted of students trained on an equine leg simulator. This division followed the pre-existent grouping for other practical activities during the semester.

Each group participated in the practical session for approximately two hours, including the time allocated for the OSCE.

Each group was further divided into smaller sub-groups, to achieve the lowest possible student-to-trainer ratio.

[2] American College of Veterinary Sports Medicine and Rehabilitation®

Immediately after the training session, students from both groups were evaluated using an OSCE structured on the procedure. All the OSCEs were evaluated by the same expert clinician using a cadaveric leg.

After the OSCE session, each student received two QR codes to complete two questionnaires: one about their satisfaction with the activity, the other on the perceived efficacy of the proposed training.

The study design, repeated for each cohort, is summarised in Figure 1.

The data collected were anonymized by assigning a numeric identification to each participant.

The purpose of the study was explained to all participants, and the informed consent was obtained and collected from each one of them.

Ethical authorization was obtained from the local Ethical Committee (University of Turin, Prot. N. 0665430, 12/27/2023).

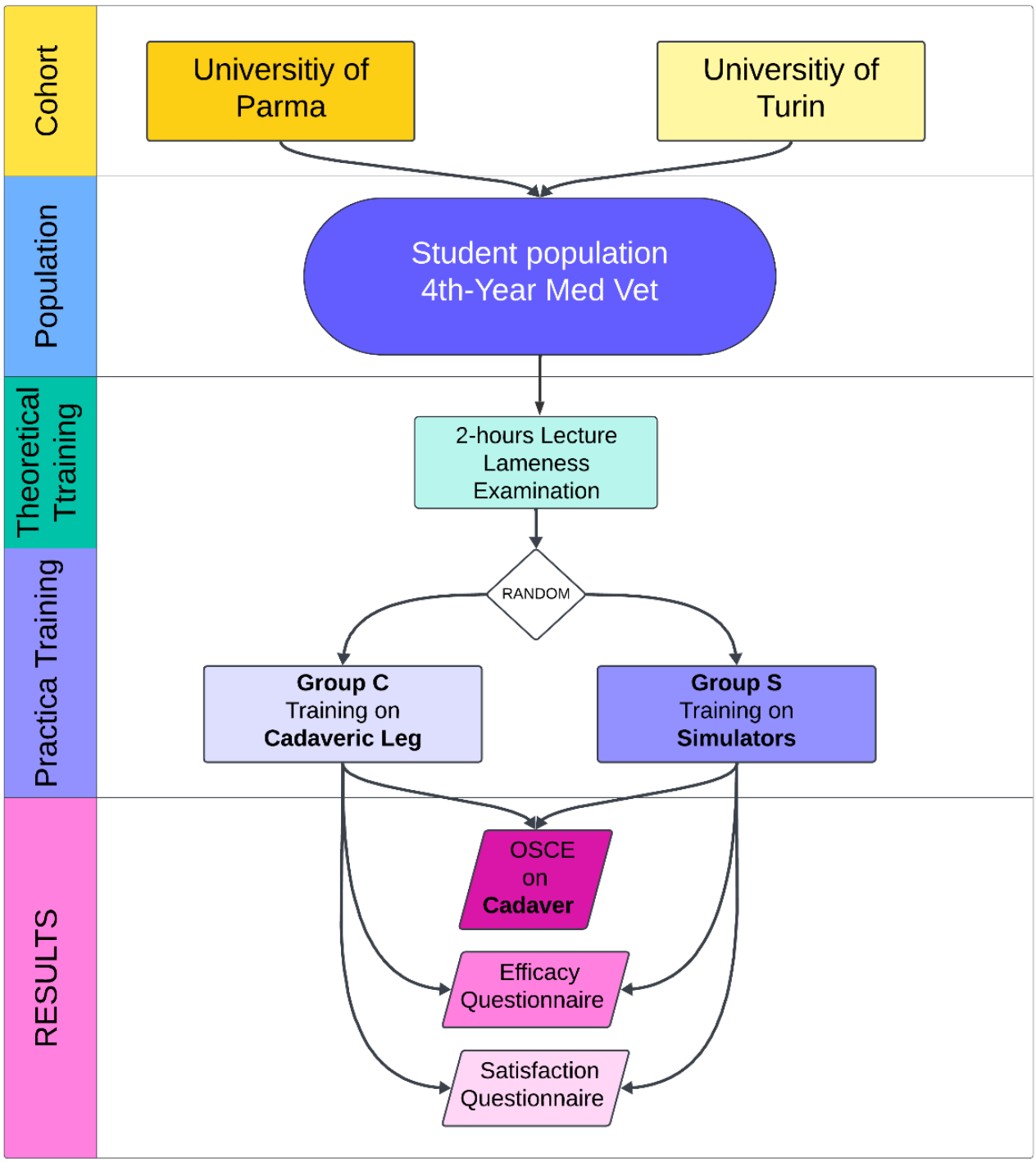


Figure 1 – Flowchart on the study design.

Lecture

The lectures were part of the curricular courses in Surgical Pathology (University of Turin) and Semiotics and Veterinary Surgery (University of Parma), offered during the fourth year of the Veterinary Medicine programs at both institutions, prior to the practical training sessions.

The lecture format consisted of a frontal presentation delivered by an associate professor at the University of Turin, ACVSMR diplomate, which gave the same lesson also to the cohort of students in Parma. The presentation was supported by slides, which included images and video materials.

The supporting materials for the presentation included 3D virtual model captures and animations from the eBook *Diagnostic Analgesia of the Equine Distal Limb*, which was released for free by the University of Georgia on Apple Books [32]. This interactive eBook combines three-dimensional anatomical animations illustrating the procedure of interest with radiographic and ultrasonographic images.

Additional supporting material for the presentation included physical 3D-printed anatomical models, commercially distributed by 3D Vet Anatomize[3]. These coloured 3D-printed models, created by Professor Christopher C. Pollitt in Australia, were based on CT and MRI scans to achieve the highest possible anatomical fidelity, making them superior tools for illustrating the complexity of equine limb anatomy to students [33].

The lectures specifically covered diagnostic anaesthesia techniques for the horse's limb, beginning with the distinction between intra-synovial and perineural blocks. Before detailing the correct procedures for patient preparation, the contraindications, as well as risks to both the operator and the patient, were thoroughly discussed.

The presentation then focused on each intra-synovial block (distal interphalangeal, proximal interphalangeal, navicular bursa, metacarpo-phalangeal, digital synovial sheath, inter-carpal, radio-carpal, tibio-tarsal, distal inter-tarsal, tarso-metatarsal,

[3] 3D Vet Anatomize © 2024 - https://3dvetanatomize.com

femoro-patellar, and femoro-tibial) and each perineural block (digital palmar/plantar nerve, abaxial sesamoid nerve, low 4-point, high 4-point, ulnar nerve, and median nerve).

For each block, the lecture presented different approaches, a brief anatomical overview, the required materials, the appropriate anaesthetic dosage, and the areas desensitized by the block.

At the conclusion of the lectures, a few minutes were dedicated to outlining the experimental project and the activities planned for the upcoming practical training sessions.

Cadaveric leg group training

One group of participants was trained on cadaveric legs. The nerve blocks proposed were:

- Digital palmar/plantar nerve block.
- Abaxial sesamoid palmar/plantar nerve block.
- Low four-point palmar/plantar nerve and palmaro/plantaro-metacarpal/metatarsal nerve block.

All the horse legs used were sourced from animals humanely euthanised at the University Teaching Hospital of Torino, with the owners' consent. The legs were stored at 4°C or -20°C, depending on the time remaining before use.

If necessary, the legs were defrosted, cleaned, and dried before the practical session and placed on the anatomy dissection tables, one leg per station.

Each station was equipped with a bottle of saline to simulate local anaesthetic, syringes (1-2.5-10-20 ml), a set of 18-20-23-25 Gauge needles and 20-23-25 Gauge butterfly needles.

Additionally, each station had laminated summary cards outlining each nerve block procedure, including anatomical images, the technique description, required materials and anaesthetic amount.

The session began with a brief introduction on the laboratory setup, and a demonstration of the three required nerve blocks.

Then, the students entered the laboratory progressively and were stationed in groups of 5-7 per table, depending on the class size and left free to practice.

Two to three instructors rotated between the stations to provide assistance and additional explanations, if required by the students, while externally supervising the execution of the required tasks and addressing any potential errors.

Once the students completed the training, they proceeded to the OSCE assessment station (described later), and the training station was taken by another student. This rotation continued until each student had practiced on the cadaveric leg and had been evaluated at the OSCE station.

Simulator group training

The second group of students trained on an equine distal limb simulator.

The simulator used was the only commercially available model at the time of the study, capable to simulate repeated nerves blocks procedure for the digital palmar, the abaxial sesamoid palmar and low four-point nerves blocks.

It consisted in a rigid, non-articulated model of the horse front leg below the carpus, constructed using 3D printing technology, except for the palmar flexor region (from the foot heel to the middle metacarpal region), which contains a replaceable silicon pad. This pad is supposed to mimics the palpation of the nerves, the haptic sensation of needle insertion, and, through an electric circuit, a led and a buzzer, provides auditory and visual feedback when the nerves are correctly targeted by a small 25-G needle.

The simulator is shown in Figure 2 with the needles positioned in the corresponding landmark for the nerves blocks of interest. The simulator is available for around 2,000 USD.

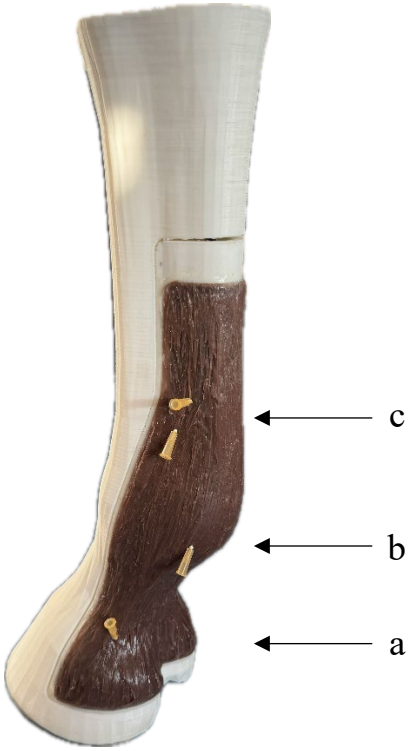


Figure 2 – Distal horse limb simulator implied in the training.
The 25G needles are positioned on the correct landmark for the nerve blocks assessed. a.: digital palmar/plantar nerves block; b.: abaxial sesamoid nerve block; c.: low four-point nerves block.

The simulator station was equipped with a complete set of syringes, needles, and butterfly needles as the cadaveric stations.

Additionally, a station featuring a 3D-printed and painted anatomical model of the equine distal limb - including skeletal structures, tendons, and ligaments - produced by 3D Vet Anatomize [33] was set up. Another station with 3D anatomy software, either Biosphera[4] or The Glass Horse[5], depending on availability in the university skills labs, was also arranged (Figure 3).

[4] Biosphera 2024 © - https://biosphera3d.com/
[5] Science In 3D, Inc. - 1151 Jeremy Dr., Watkinsville, GA 30677 - https://www.sciencein3d.com/

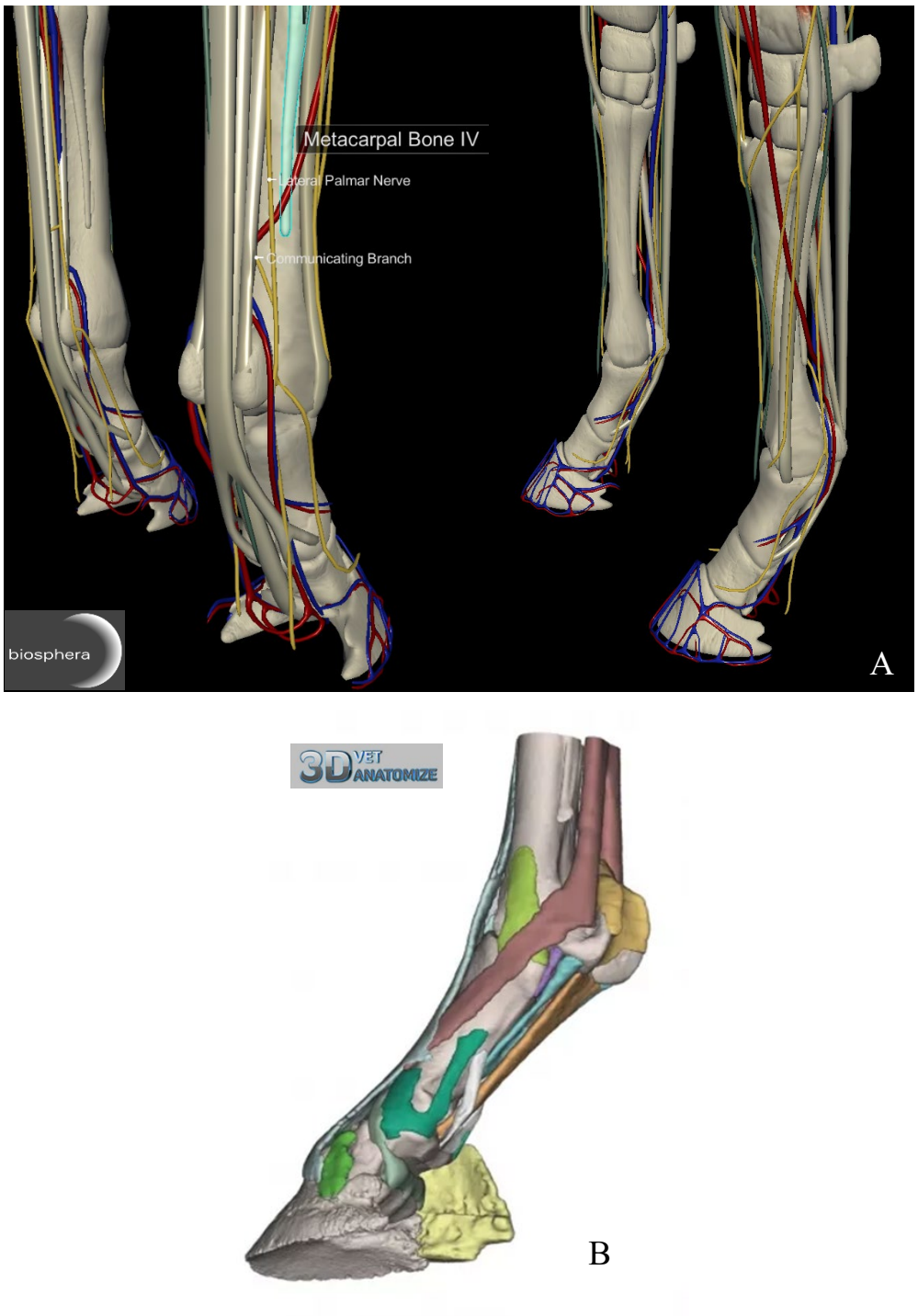


Figure 3 – Software extract and 3D printed model available to student of Group S. Panel A show captures form Biosphera Software® of the anatomy of the horse distal limb; panel B show 3D Vet Anatomize® 3D printed model of the horse distal limb.

Each simulator station also had laminated summary cards with anatomical images, technique descriptions, required material and anaesthetic amount.

Students were allowed to enter the laboratory together and were asked to distribute themselves equally among the simulation stations.

The session began with a brief introduction on what was expected from the session, the laboratory setup, the correct use of the simulator and on the following OSCE evaluation. This was followed by a demonstration of the procedure on the simulator and explanation of the anatomy software functionality.

Students were then left free to practice and rotate among the simulation stations, cooperating to solve doubts and improve their ability. An instructor externally supervised each station, providing assistance when requested, and addressing any potential errors.

The only rule was that each student had to practice each nerve block at least once on the limb simulator or continue practicing until they could correctly perform the procedure.

Once one student completed his/her training, he/she moved on to the OSCE evaluation (described below).

OSCE Station

The OSCE station aimed to provide the students with immediate feedback on their confidence in performing the nerve block procedure after the training session.

The OSCE station was set up in the pathology lab with a cadaveric leg, prepared as previously described, and a full set of needles and syringes.

The evaluation was conducted by a ACVSMR diplomate, who asked the student to autonomously perform a randomly selected nerve block and evaluated four criteria:

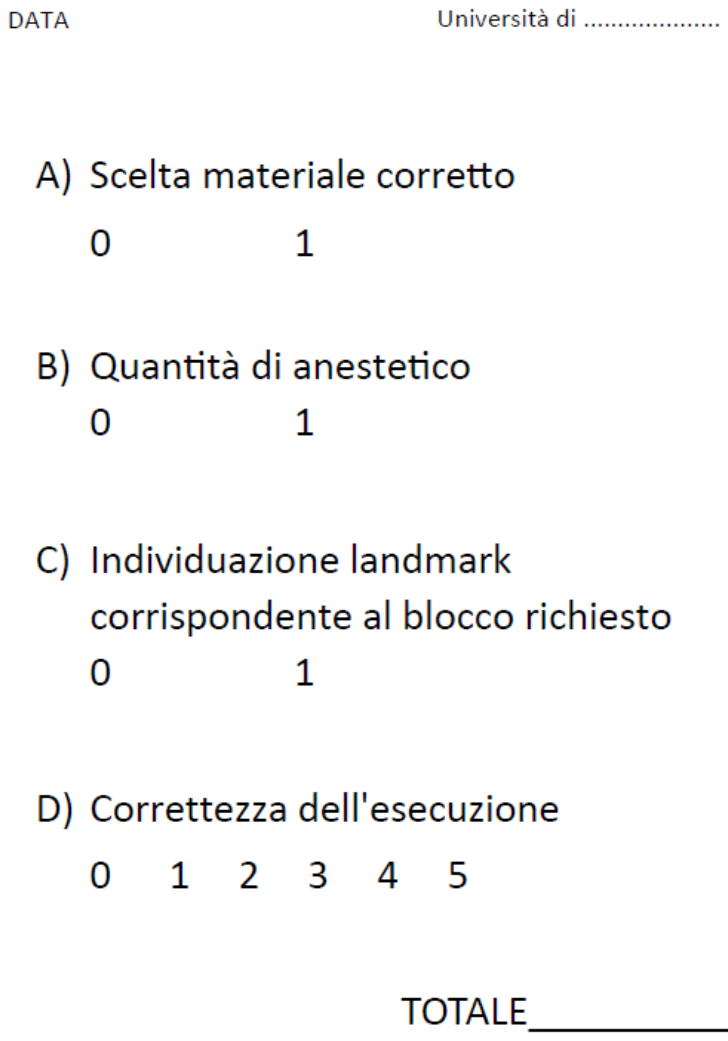
- cA. Choice of materials (needle and syringe)
- cB. Amount of local anaesthetic
- cC. Identification of the landmark for the targeted nerve block
- cD. General proficiency in manual skills and technique

The examiner, the same ACVSMR diplomate for the entire study, was provided with a form for evaluating the criteria: criteria A, B, and C were marked as Passed/Failed (1

or 0 points, respectively), criterion D was graded on a scale from 0 to 5 (were 0 indicated insufficient skill or incorrect technique and 5 indicated optimal skill and correct technique) (Figure 4).

The overall score was calculated by summing the points for all four criteria, with a total score ranging from 0 to 8. The result was briefly discussed with the student at the end of the evaluation to suggest corrective actions and areas for improvement, if necessary.

The back of the form contained the student's numeric identification to anonymize the data and match the OSCE score with the student's questionnaire responses.


DATA Università di .....

A) Scelta materiale corretto

0 1

B) Quantità di anestetico

0 1

C) Individuazione landmark corrispondente al blocco richiesto

0 1

D) Correttezza dell'esecuzione

0 1 2 3 4 5

TOTALE_____

Figure 4 – Model for the OSCE form filled by the examiner for each student.

Questionnaires

Two questionnaires were provided to students via QR codes after the practical session, with the request to fill in them online (from smartphone, PC, or tablet platform) within the next 24 hours.

The first questionnaire addressed the students' satisfaction on the practical activity (7 items), the second assessed their perceived efficacy of the training (5 items).

The questionnaires were proposed in Italian (students' native language). A 5-points Likert scale was selected to allow participants for providing nuanced responses to attitudinal and perception-based questions, offering a range of agreement levels rather than binary choices. So, the items required the participants to rate their agreement with each statement on a 5-point scale: 0 - completely disagree, 1 - disagree, 2 - neutral, 3 - agree, 4 - strongly agree. An optional free comment section was included at the end of each questionnaire.

All data were collected anonymously by requesting that participants only input their identification number.

The items were adapted from those used by Rajaure et al. in 2023 [34] for a similarly designed study.

This questionnaire was considered the most pragmatic way to measure a non-quantifiable parameter such as the student satisfaction and the perceived efficacy of the two practical training methods used in this study [35].

DATA COLLECTION AND STATISTICAL ANALYSIS

The data collected from the OSCE evaluation and questionnaires from the two Universities were entered into a Microsoft® Excel®[6] spreadsheet and analysed by STATA v.15.1[7].

For each participant, the following data were recorded:

- The scores for criterion assessed with the OSCE
- The overall OSCE score
- The score for each question in the satisfaction and efficacy questionnaires

Using these data, the following parameters were calculated for each participant:

- The cumulative OSCE score, obtained by summing the four key points assessed with OSCE
- The cumulative scores of the satisfaction and efficacy questionnaires, obtained by summing the scores of their single questionnaire items.

Data distribution was deemed as non-parametric based on their non-continuous scale. Data were presented as median (25^{th}, 75^{th} percentile) or as raw counts and percentage. Differences between two groups were assessed statistically using nonparametric tests, namely Mann-Whitney test to compare medians, and chi-squared test to compare proportions. Linear regression analysis was used to adjust for multiple covariates when deemed necessary.

Sensitivity analysis of the data

The sensitivity analysis was conducted *a posteriori* for testing the power of the study for results obtained between cohorts (University of Parma vs. University of Turin) and between groups (University of Parma: Group S vs. Group C; University of Turin: Group S vs. Group C).

[6] Microsoft Corporation. Microsoft® Excel® for Microsoft 365 MSO (v2409)
[7] StataCorp. 2023. Stata Statistical Software: Release 18. College Station, TX: StataCorp LLC.

The effect size quantifies the difference between two groups. While a p-value indicates whether there is a statistically significant difference, an effect size reveals the magnitude of that difference.

There are various methods to measure effect size, depending on the type of analysis being conducted. In our case, since our primary objective was to evaluate the difference between groups, the most suitable measure is the Standardized Mean Difference. The most widely used formula is known as Cohen's d, which is calculated as:

$$Cohen's\ d = \frac{(\bar{x}_1 - \bar{x}_2)}{s}$$

where $\bar{x}_1$ and $\bar{x}_2$ are the sample means of Group 1 and Group 2, respectively, and s is the standard deviation of the population from which the two groups were taken.

Using this formula, the effect size (d) is easy to interpret:

- A d of 1 indicates that the means of the two groups differ by one standard deviation
- A d of 2 indicates a difference of two standard deviations between group means
- A d of 2.5 means the group means differ by 2.5 standard deviations, and so on.

Another way to understand effect size is by comparing average scores between groups. For instance, an effect size of 0.3 means that the average score of an individual in Group 2 is 0.3 standard deviations above that of an individual in Group 1, placing them above 62% of individuals in Group 1.

In general, a larger effect size indicates a more substantial difference between the average scores of individuals in each group.

Conventionally:

- A d of 0.2 or less is considered a small effect size
- A d around 0.5 is regarded as a medium effect size
- A d of 0.8 or more is considered a large effect size. [36]

University of Parma cohort

A sensitivity analysis was conducted *a posteriori* using GPower software for testing the power of the study for results obtained between the groups of the University of Parma cohort. This allowed determining the minimum detectable effect size given the sample size available at the end of the experimental session (n=33), significance level ($\alpha = 0.05$), and desired power ($1-\beta = 0.80$). This analysis helps to ensure the robustness of the study design by evaluating the likelihood of Type I errors.

Table 1 details the options used in GPower software to conduct the analysis, results are presented below and represented in Figure 5.

Test family	t tests
Statistical test	Means: Wilcoxon-Mann-Whitney test (two groups)
Type of power analysis	Sensitivity: Compute required effect size – given α, power, and sample size
Input Parameters	
Tail(s)	Two
Parent distribution	Normal
α err prob	0.05
Power (1-β err prob)	0.80
Sample size group 1	13
Sample size group 2	20

Table 1 – GPower software options setup for sensitivity analysis of data for comparison between Groups within the University of Parma Cohort. Parameters provided for analysis reproducibility.

University of Turin cohort

A sensitivity analysis was conducted *a posteriori* using GPower software for testing the power of the study for results obtained between the groups of the University of Turin cohort. This allowed determining the minimum detectable effect size given the sample size available at the end of the experimental session (n=37), significance level ($\alpha = 0.05$), and desired power ($1-\beta = 0.80$). This analysis helps to ensure the robustness of the study design by evaluating the likelihood of Type I errors.

Table 2 details the options used in GPower software to conduct the analysis, results are presented below and represented in Figure 6.

Test family	t tests
Statistical test	Means: Wilcoxon-Mann-Whitney test (two groups)
Type of power analysis	Sensitivity: Compute required effect size – given α, power, and sample size
Input Parameters	
Tail(s)	Two
Parent distribution	Normal
α err prob	0.05
Power (1-β err prob)	0.80
Sample size group 1	17
Sample size group 2	20

Table 2 - GPower software options setup for sensitivity analysis of data for comparison between Groups within the University of Turin Cohort. Parameters provided for analysis reproducibility.

University of Parma cohort vs. University of Turin cohort

A sensitivity analysis was conducted using GPower software *a posteriori* for comparison of results between cohorts. This allowed to determine the minimum detectable effect size given the sample size available at the end of the experimental session (n=70), significance level (α = 0.05), and desired power (1-β = 0.80). This analysis helps to ensure the robustness of the study design by evaluating the likelihood of Type I errors.

Table 3 details the options used in GPower software to conduct the analysis, results are presented below and represented in Figure 7.

Test family	t tests
Statistical test	Means: Wilcoxon-Mann-Whitney test (two groups)
Type of power analysis	Sensitivity: Compute required effect size – given α, power, and sample size
Input Parameters	
Tail(s)	Two
Parent distribution	Normal
α err prob	0.05
Power (1-β err prob)	0.80
Sample size group 1	33
Sample size group 2	37

Table 3 – GPower software options setup for sensitivity analysis of data for comparison between cohorts. Parameters provided for analysis reproducibility.

RESULTS

University of Parma cohort

Description of groups

A total of 41 students from the cohort of the University of Parma participated in this trial, whose demographic data were not collected. Twenty-one of them were assigned to Group C (those trained on the cadaveric leg), and the remaining 20 were assigned to Group S (those trained on the horse limb simulator). Group distribution is summarised in Table 4.

	UNIPR		
	GROUP C	**GROUP S**	**TOT**
TOTAL [n]	21	20	41
OSCE [n (%)]	20 (95.2)	13 (65)	33 (80.5)
EFFICACY Q. [n (%)]	14 (66.7)	13 (65)	27 (65.9)
SATISFACTION Q. [n (%)]	14 (66.7)	15 (75)	29 (69)

Table 4 – Group distribution and participation rate for the University of Parma cohort. Data are presented as number of student (participation rate, %).

OSCE

In the University of Parma cohort, a total of 33 students out of 41 participated into the activity proposed at the OSCE station, accounting for an overall 80.5% participation rate. Eight students did not show up for the OSCE examination. When expressed by group, a total of 20 out of 21 students of Group C participated into the activity proposed at the OSCE station versus a total of 13 out of 20 students of Group S, corresponding, to 95.2% vs. 65% participation rate, respectively ($p=0.01$, chi-squared test). Participation rate is summarised in Table 4.

The overall success rate was 91% for criterion A, 71% for criterion B and 55% for criterion C. The median score for criterion D was 3 (3, 5). Detailed results are presented in Table 5.

Considering the group division, students in Group C obtained statistically significant higher total scores in the OSCE compared to students in Group S ($p= 0.006$, Mann-Whitney).

The scores obtained by the students in Groups S and C were similar for the criteria A, B, and C ($p> 0.05$, Chi-squared test).

In detail, for Group C the success rate was 95% for criterion A, 80% for criterion B, and 60% for criterion C, while, for Group S the success rate was 85% for criterion A, 69% for criterion B, and 46% for criterion C.

A significant difference in the score gained in the criteria was only observed for criterion D (0.004, Mann-Whitney test). In detail, the median response for this criterion for the Group C was 5 (3.5, 5), while for the Group S the median response was 3 (3, 3). OSCE scores obtained by students in each group studied are presented in Table 5.

OSCE CRITERIA	DETAIL	SCORING SYSTEM	GROUP C+S	GROUP C	GROUP S	P-VAL
cA	Choice of materials	Passed (%) vs. failed (%)	30 (91) vs. 3 (9)	19 (95) vs. 1 (5)	11 (85) vs. 2 (15)	0.311
cB	Amount of local anaesthetic	Passed (%) vs. failed (%)	25 (76) vs. 8 (24)	16 (80) vs. 4 (20)	9 (69) vs. 4 (31)	0.481
cC	Identification of the landmark for the targeted nerve block	Passed (%) vs. failed (%)	18 (55) vs. 15 (45)	12 (60) vs. 8 (40)	6 (46) vs. 7 (54)	0.435
cD	**General proficiency in manual skills and techniques**	**5-point score (5 = max)**	**3 (3, 5)**	**5 (3.5, 5)**	**3 (3, 3)**	**0.004**

Table 5 – OSCE Results for University of Parma cohort.
For cA, cB, and cC, student assessment data are expressed as Number of students that passed vs. failed the criteria. For cD, data are presented as median (25^{th}, 75^{th} percentile) of data collected using a 5-points scoring scale (max = 5). P-value is calculated with χ^2 for cA, cB and cC, and with Mann-Whitney for cD. Bold indicates statistically significant difference between the groups.

Efficacy Questionnaire

A total of 27 out of 41 students filled the Efficacy Questionnaire, accounting for an overall 65.9% completion rate. When expressed by group, a total of 14 out of 21 students of Group C completed the Efficacy Questionnaire, versus a total of 13 out of 20 students of Group S, corresponding to 66.7% vs. 65% participation rate, respectively (p=0.91, chi-squared test). Participation rate is summarised in Table 4.

Concerning the outcome in the Efficacy Questionnaire, there was no statistically significant difference in the total score obtained between the groups (p= 0.203, Mann-Whitney).

The first item of the efficacy questionnaire was “I am confident that I am obtaining the required knowledge from this clinical skill session to perform the proposed tasks in the clinical practice”. The median response to this item was 4 (3, 4). There was no

statistically significant difference in the score obtained in the groups (p= 0.190, Mann-Whitney).

The second item in the Efficacy Questionnaire was “I am confident that I am developing the required skills from the clinical skill session to perform the tasks in a clinical setting”. The median response was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.067, Mann-Whitney).

The third item in the Efficacy Questionnaire was “I am certain that I can accomplish my intended learning objectives for these sessions”. The median response was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.526, Mann-Whitney).

The fourth item in the Efficacy Questionnaire was “I am confident that, after the session, I am mastering the content of the simulation activity that my facilitators presented to me”. The median response was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.391, Mann-Whitney).

The fifth item in the Efficacy Questionnaire was “I am confident that the clinical skills sessions covered all the necessary content mentioned in the curriculum”. The median response was 4 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.394, Mann-Whitney).

Two students released a free comment, both from Group C. The comments were “The exercise is extremely useful; however, as far as I am concerned, I cannot say that I would be able to replicate the practice on a living subject, due to a personal fear rather than because of its ineffectiveness” and “Useful and formative exercise, with very professional staff”.

The efficacy questionnaire scores obtained by the two groups studied are summarized in Table 6.

EFFICACY ITEM	DETAIL	GROUP C+S	GROUP C	GROUP S	P-VALUE
i1	I am obtaining required knowledge to perform the proposed tasks in the clinical practice	4 (3, 4)	4 (3, 4)	3 (3, 4)	0.190
i2	I am developing required skills to perform the tasks in a clinical practice	3 (3, 4)	4 (3, 4)	3 (3, 3)	0.067
i3	I can accomplish my intended learning objectives for these sessions	3 (3, 4)	3 (3, 4)	3 (3, 4)	0.526
i4	I am mastering the content of the simulation activity	3 (3, 4)	3 (3, 4)	3 (2, 4)	0.392
i5	The clinical skills sessions covered all the necessary content mentioned in the curriculum	4 (3, 4)	4 (3, 4)	3 (3, 4)	0.394
RESPONDERS [n(%)]		27 (65.9)	14 (66.7)	13 (65)	

Table 6 – Efficacy Questionnaire Results for University of Parma cohort.
Data are presented as median (25th, 75th percentile) of data collected using a 5-points Likert scale (max = 5). P-value is calculated with Mann-Whitney. The last row reports the number of responders and, in brackets, participation rate expressed as a percentage.

Satisfaction Questionnaire

A total of 29 out of 42 students filled the satisfaction questionnaire, accounting for an overall 69% completion rate. When expressed by group, a total of 14 out of 21 students of Group C completed the Satisfaction Questionnaire, versus a total of 15 out of 20 students of Group S (corresponding to 66.7% vs. 75% participation rate, respectively; p=0.56, chi-squared test). Participation rate is summarised in Table 4.

As for the efficacy questionnaire, there was no statistically significant difference in the total score obtained in the groups (p= 0.912, Mann-Whitney).

The first item of the satisfaction questionnaire was “The integration of simulation with manikin/the training session method for clinical skills teaching was effective in achieving the learning objectives”. The median response was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.595, Mann-Whitney).

The second item of the Satisfaction Questionnaire was “The integration of simulation with manikin/training session was well organized in terms of scheduling and planning”. The median response was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.333, Mann-Whitney).

The third item of the Satisfaction Questionnaire was “The facilitators gave me clear instructions of what is expected from me during this session”. The median response was 4 (4, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.748, Mann-Whitney).

The fourth item of the Satisfaction Questionnaire was “The integration of simulation with manikin/training session method for clinical skills teaching were motivating me to learn”. The median response was 4 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.742, Mann-Whitney).

The fifth item of the Satisfaction Questionnaire was “Sufficient guidance was given to me by the facilitator before I performed on simulation”. The median response was 4 (4, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.620, Mann-Whitney).

The sixth item of the Satisfaction Questionnaire was “The way my facilitators conducted the simulation was suitable to the way I learn”. The median response was 4 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.892, Mann-Whitney).

The seventh item of the Satisfaction Questionnaire was “The integration of simulation with manikin/training session method for the clinical skills teaching helped me to link theory to practice”. The median response was 4 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.403, Mann-Whitney).

Two students released a free comment, one from Group C and one from Group S. The comments were “It was very useful to immediately put into practice what I had previously only seen in theory” and “The only difficulty I encountered with the first

approach to the mannequin was the correct palpation of the landmarks (in my opinion, incomparable to what is achieved on a living subject)".

The satisfaction questionnaire scores obtained by the two groups studied are summarized in Table 7.

SATISFACTION ITEM	DETAIL	GROUP C+S	GROUP C	GROUP S	P-VALUE
i1	The training session was effective in achieving the learning objectives	3 (3, 4)	3.5 (3, 4)	3 (3, 4)	0.595
i2	The training session was well organized in terms of scheduling and planning	3 (3, 4)	3 (2, 4)	4 (3, 4)	0.333
i3	The facilitators gave me clear instructions of what is expected from me during this session	4 (4, 4)	4 (4, 4)	4 (3, 4)	0.748
i4	The training session method for clinical skills teaching were motivating me to learn	4 (3, 4)	4 (3, 4)	4 (3, 4)	0.742
i5	Sufficient guidance was given to me by the facilitator before I performed on simulation	4 (4, 4)	4 (3, 4)	4 (4, 4)	0.620
i6	The way my facilitators conducted the simulation was suitable to the way I learn	4 (3, 4)	4 (3, 4)	4 (3, 4)	0.893
i7	The training session method for the clinical skills teaching helped me to link theory to practice	4 (3, 4)	4 (3, 4)	4 (3, 4)	0.403
RESPONDERS [n(%)]		29 (69)	14 (66.7)	15 (75)	

Table 7 – Satisfaction Questionnaire Results for University of Parma cohort.
Data are presented as median (25th, 75th percentile) of data collected using a 5-points Likert scale (max = 5). P-value is calculated with Mann-Whitney. The last row reports the number of responders and, in brackets, participation rate expressed as a percentage.

Other results

No significant correlation was detected between the score obtained in the OSCE and the Satisfaction and Efficacy questionnaire results (p=0.77, r=0.07; and p=0.43, r=0.19, respectively, Spearman test). However, a significant correlation was observed between the Satisfaction and Efficacy questionnaire results (p=0.0003, r=0.74, Spearman test). Data are summarised in Table 14.

Sensitivity analysis of the data

The sensitivity analysis run with the data from the groups of the University of Parma cohort indicated that, with a sample size of 33 participants and an alpha level of 0.05, the study is powered (80%) for observing a difference between the performances of students from the two groups studied which corresponds to an effect size d of 1.05 (extremely large) (Figure 5). For smaller effects, there is a high risk of Type I error.

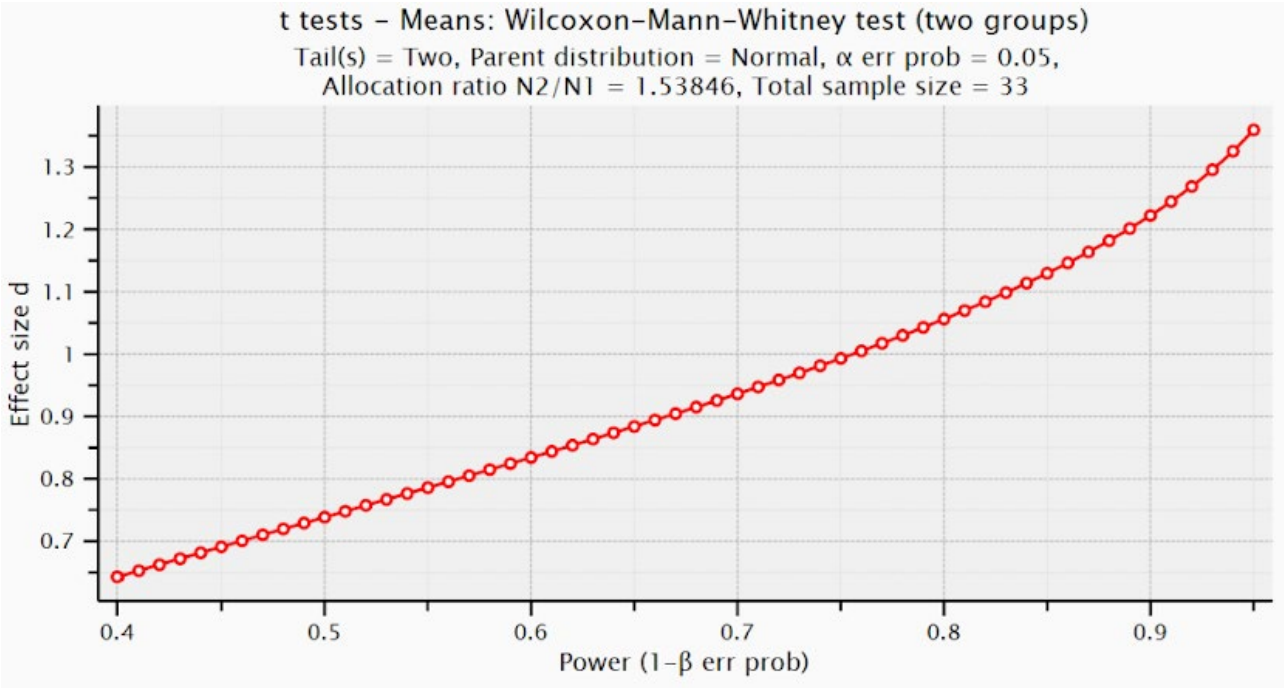


Figure 5 - Sensitivity Analysis for Wilcoxon–Mann–Whitney Test (Two Groups) Using GPower Software.
This graph depicts the relationship between statistical power (1-β error probability) and the minimum detectable effect size (Cohen's d) for a Wilcoxon–Mann–Whitney test with two groups defined as N1=13 and N2=20, corresponding to the simple sizes of students performing OSCE in Group S and Group C, respectively, at University of Parma cohort. The sensitivity analysis was conducted considering a two-tailed analysis with a normal parent distribution, and an α error probability of 0.05. For a 80% power of the study, the effect size d = 1.05 (extremely large).

University of Turin cohort

Description of groups

A total of 37 students from the cohort of the University of Turin participated in this trial, whose demographic data were not collected. Twenty of them were assigned to Group C (those trained on the cadaveric leg), while the remaining 17 were assigned to Group S (those trained on the horse limb simulator). Group division is summarised in Table 8.

	UNITO		
	GROUP C	**GROUP S**	**TOT**
TOTAL [n]	20	17	37
OSCE [n (%)]	20 (100)	17 (100)	37 (100)
EFFICACY Q. [n (%)]	19 (95)	14 (82.4)	33 (89.2)
SATISFACTION Q. [n (%)]	16 (80)	15 (88.2)	31 (84)

Table 8 - Group distribution and participation rate for the University of Turin cohort. Data are presented as number of student (participation rate, %).

OSCE

In the University of Turin cohort, a total of 37 students out of 37 participated into the activity proposed at the OSCE station, accounting for an overall 100% participation rate. Participation rate is summarised in Table 8.

The overall success rate was 92% for criterion A, 95% for criterion B and 95% for criterion C. The median score for criterion D was 4 (3, 5). Detailed results are presented in Table 9.

Considering the groups division, no statistically significant difference in the total score obtained in the OSCE between Group C and Group S ($p= 0.9$, Mann-Whitney) was observed.

The scores obtained by the students in Groups S and C were similar for the criteria A, B, C and D ($p > 0.05$, Chi-squared test).

OSCE score obtained by students in each group studied are presented in Table 9.

In detail, for Group C the success rate was 95% for criterion A, 90% for criterion B, and 95% for criterion C, while, for Group S the success rate was 88% for criterion A, 100% for criterion B, and 94% for criterion C.

OSCE CRITERIA	DETAIL	SCORING SYSTEM	GROUP C+S	GROUP C	GROUP S	P-VAL
cA	Choice of materials	Passed (%) vs. failed (%)	34 (92) vs. 3 (8)	19 (95) vs. 1 (5)	15 (88) vs. 2 (12)	0.452
cB	Amount of local anaesthetic	Passed (%) vs. failed (%)	35 (95) vs. 2 (5)	18 (90) vs. 2 (10)	17 (100) vs. 0 (0)	0.180
cC	Identification of the landmark for the targeted nerve block	Passed (%) vs. failed (%)	35 (95) vs. 2 (5)	19 (95) vs. 1 (5)	16 (94) vs. 1 (6)	0.906
cD	General proficiency in manual skills and techniques	5-point score (5 = max)	4 (3, 5)	4 (3, 5)	4 (3, 5)	0.987

Table 9 – OSCE Results for University of Turin cohort.
For cA, cB, and cC, student assessment data are expressed as Number of students that passed vs. failed the criteria. For cD, data are presented as median (25^{th}, 75^{th} percentile) of data collected using a 5-points scoring scale (max = 5). P-value is calculated with χ^2 for cA, cB and cC, and with Mann-Whitney for cD.

Efficacy Questionnaire

A total of 33 out of 37 students filled the Efficacy Questionnaire, accounting for an overall 89.2% completion rate. When expressed by group, a total of 19 out of 20

students of Group C completed the Efficacy Questionnaire, versus a total of 14 out of 17 students of Group S, corresponding to 95% vs. 82.4% participation rate, respectively (p=0.22, chi-squared test). Participation rate is summarised in Table 8.

Concerning the outcome in the Efficacy Questionnaire, there was no statistically significant difference in the total score obtained in the groups (p= 0.312, Mann-Whitney).

The first item of the efficacy questionnaire was “I am confident that I am obtaining the required knowledge from this clinical skill session to perform the proposed tasks in the clinical practice”. The median response for this item was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.984, Mann-Whitney).

The second item in the Efficacy Questionnaire was “I am confident that I am developing the required skills from the clinical skill session to perform the tasks in a clinical practice”. The median response was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.211, Mann-Whitney).

The third item in the Efficacy Questionnaire was “I am certain that I can accomplish my intended learning objectives for these sessions”. The median response was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.195, Mann-Whitney).

The fourth item in the Efficacy Questionnaire was “I am confident that, after the session, I am mastering the content of the simulation activity that my facilitators presented to me”. The median response was 3 (2, 3). There was no statistically significant difference in the score obtained in the groups (p= 0.922, Mann-Whitney).

The fifth item in the Efficacy Questionnaire was “I am confident that the clinical skills sessions covered all the necessary content mentioned in the curriculum”. The median response was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.234, Mann-Whitney).

Two students released a free comment, one from each group. The comments were "Poor clarity in the initial explanation given by the teacher" from the student from Group C, and "Unrealistic simulator for reference points" from the student from Group S.

The efficacy questionnaire scores obtained by the two groups studied are summarized in Table 10.

EFFICACY ITEM	DETAIL	GROUP C+S	GROUP C	GROUP S	P-VALUE
i1	I am obtaining required knowledge to perform the proposed tasks in the clinical practice	3 (3, 4)	3 (3, 4)	3.5 (3, 4)	0.984
i2	I am developing required skills to perform the tasks in a clinical practice	3 (3, 4)	3 (3, 4)	3 (3, 4)	0.211
i3	I can accomplish my intended learning objectives for these sessions	3 (3, 4)	3 (3, 4)	3 (2, 4)	0.195
i4	I am mastering the content of the simulation activity	3 (2, 3)	3 (2, 3)	3 (2, 4)	0.922
i5	The clinical skills sessions covered all the necessary content mentioned in the curriculum	3 (3, 4)	3 (3, 4)	3 (2, 3)	0.234
RESPONDERS [n(%)]		33 (89.2)	19 (95)	14 (82.4)	

Table 10 – Efficacy Questionnaire Results for University of Turin cohort.
Data are presented as median (25th, 75th percentile) of data collected using a 5-points Likert scale (max = 5). P-value is calculated with Mann-Whitney. The last row reports the number of responders and, in brackets, participation rate expressed as a percentage.

Satisfaction Questionnaire

A total of 31 out of 37 students filled the satisfaction questionnaire, accounting for an overall 84% completion rate. When expressed by group, a total of 16 out of 20 students of Group C completed the Satisfaction Questionnaire, versus 15 out of 17 students of Group S (corresponding to 80% vs. 88.2% participation rate, respectively; p=0.5, chi-squared test). Participation rate is summarised in Table 8.

As for the efficacy questionnaire, there was no statistically significant difference in the total score obtained in the groups (p= 0.474, Mann-Whitney).

The first item of the satisfaction questionnaire was “The integration of simulation with manikin/the training session method for clinical skills teaching was effective in achieving the learning objectives”. The median response was 3 (3, 4). A significant difference in the response to this question was observed between groups (0.005, Mann-Whitney test). In detail, the median response for the Group C was 4 (3, 4), and for the Group S the median response was 3 (2, 3).

The second item of the Satisfaction Questionnaire was “The integration of simulation with manikin/training session was well organized in terms of scheduling and planning”. The median response was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.079, Mann-Whitney).

The third item of the Satisfaction Questionnaire was “The facilitators gave me clear instructions of what is expected from me during this session”. The median response was 4 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.873, Mann-Whitney).

The fourth item of the Satisfaction Questionnaire was “The integration of simulation with manikin/training session method for clinical skills teaching were motivating me to learn”. The median response was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.242, Mann-Whitney).

The fifth item of the Satisfaction Questionnaire was “Sufficient guidance was given to me by the facilitator before I performed on simulation”. The median response was 3 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.115, Mann-Whitney).

The sixth item of the Satisfaction Questionnaire was “The way my facilitators conducted the simulation was suitable to the way I learn”. The median response was 4 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.903, Mann-Whitney).

The seventh item of the Satisfaction Questionnaire was “The integration of simulation with manikin/training session method for the clinical skills teaching helped me to link theory to practice”. The median response was 4 (3, 4). There was no statistically significant difference in the score obtained in the groups (p= 0.828, Mann-Whitney).

Two students released a free comment, both from Group S. The comments were “The use of the mannequin did not realistically reflect the anatomical structures involved during the test. I did not encounter any difficulties in performing the anaesthesia procedures on the mannequin, whereas I was unable to perform them correctly on the horse leg” and “It is hard to palpate the anatomical structures on the mannequin”.

The satisfaction questionnaire scores obtained by the two groups studied are summarized in Table 11.

SATISFACTION ITEM	DETAIL	GROUP C+S	GROUP C	GROUP S	P-VALUE
i1	**The training session was effective in achieving the learning objectives**	**3 (3, 4)**	**4 (3, 4)**	**3 (2, 3)**	**0.005**
i2	The training session was well organized in terms of scheduling and planning	3 (3, 4)	3 (3, 4)	3 (3, 3)	0.079
i3	The facilitators gave me clear instructions of what is expected from me during this session	4 (3, 4)	4 (3, 4)	4 (3, 4)	0.873
i4	The training session method for clinical skills teaching were motivating me to learn	3 (3, 4)	3.5 (3, 4)	3 (3, 4)	0.242
i5	Sufficient guidance was given to me by the facilitator before I performed on simulation	3 (3, 4)	3 (3, 4)	3 (3, 4)	0.115
i6	The way my facilitators conducted the simulation was suitable to the way I learn	4 (3, 4)	4 (3, 4)	4 (3, 4)	0.903
i7	The training session method for the clinical skills teaching helped me to link theory to practice	4 (3, 4)	3.5 (3, 4)	4 (3, 4)	0.828
RESPONDERS [n(%)]		31 (84)	16 (80)	15 (88.2)	

Table 11 – Satisfaction Questionnaire Results for University of Turin cohort. Data are presented as median (25th, 75th percentile) of data collected using a 5-points Likert scale (max = 5). P-value is calculated with Mann-Whitney. Bold indicates statistically significative difference between the groups. The last row reports the number of responders and, in brackets, participation rate expressed as a percentage.

Other results

No significant correlation was detected between the score obtained in the OSCE and the Satisfaction and Efficacy questionnaire results ($p=0.24$, $r=0.23$; and $p=0.07$, $r=0.35$, respectively, Spearman test). However, a significant correlation was observed between the Satisfaction and Efficacy questionnaire results ($p<0.0001$, $r=0.75$, Spearman test). Data are summarised in Table 14.

Sensitivity analysis of the data

The sensitivity analysis run with data from the groups of the University of Turin cohort indicated that, with a sample size of 37 participants and an alpha level of 0.05, the study is powered (80%) for observing a difference between the performances of students from the two groups studied which corresponds to an effect size d of 0.98 (extremely large) (Figure 6). For smaller effects, there is a high risk of Type I error.

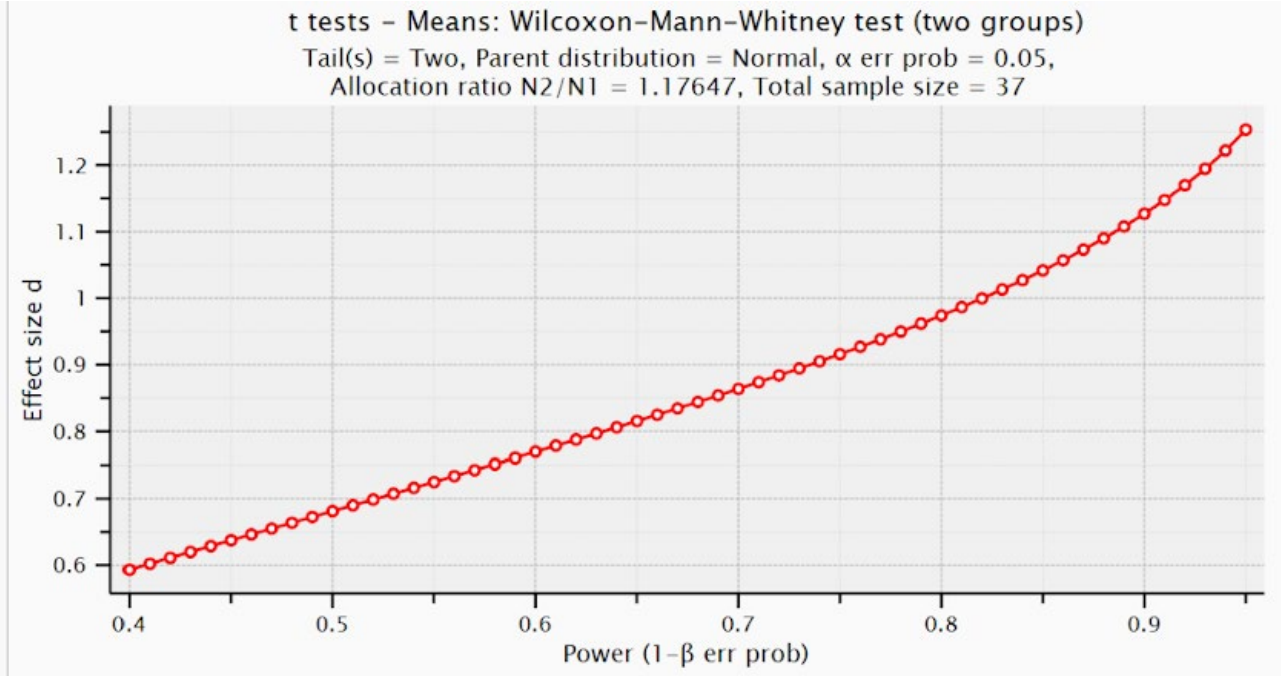


Figure 6 - Sensitivity Analysis for Wilcoxon–Mann–Whitney Test (Two Groups) Using GPower Software.
This graph depicts the relationship between statistical power (1-β error probability) and the minimum detectable effect size (Cohen's d) for a Wilcoxon–Mann–Whitney test with two groups defined as N1=17 and N2=20, corresponding to the simple sizes of students performing OSCE in Group S and Group C, respectively, at University of Turin cohort. The sensitivity analysis was conducted considering a two-tailed analysis with a normal parent distribution, and an α error probability of 0.05. For an 80% power of the study, the effect size d = 0.98 (extremely large).

University of Turin cohort vs. University of Parma cohort

Comparison between the two Departments was performed using linear regression model with robust error calculation, adjusted by the Group (C vs. S) for the following parameters: overall OSCE score, singular OSCE criteria, overall satisfaction questionnaire and overall efficacy questionnaire.

The results indicate an overall better performance of University of Turin students vs. University of Parma students for criteria B (β coefficient (95% C.I.) = 0.19 (0.01, 0.36); p = 0.03) and C of the OSCE (β coefficient (95% C.I.) = 0.40 (0.21, 0.60); p < 0.0001), while no difference was observed for criteria A and D. Overall, no significant difference was observed for the overall OSCE scores between the two cohorts studied (β coefficient (95% C.I.) = 0.79 (-0.06, 1.64); p = 0.07).

Overall results of Satisfaction and Efficacy questionnaires were similar in the two cohorts studied.

Results of those analysis are summarised in Table 12.

	β coefficient	**95% C.I.**	**p value**
OSCE	0.79	-0.06, 1.64	0.068
cA	0.02	-0.12, 0.15	0.823
cB	**0.19**	**0.01, 0.36**	**0.034**
cC	**0.40**	**0.21, 0.60**	**<0.0001**
cD	0.18	-0.47, 0.84	0.578
Efficacy Questionnaire	-1.14	-2.65, 0.38	0.139
Satisfaction Questionnaire	-1.45	-3.15, 0.25	0.094

Table 12 – Effect size of the cohort (University of Turin vs. Parma) adjusted for Group effect (C vs. S).
Bold indicates statistically significant effects.

Sensitivity analysis of the data

The sensitivity analysis on the comparison of the results of the two Universities indicated that, with a sample size of 70 participants and an alpha level of 0.05, the study is powered (80%) for observing a difference between the performances of students from the two Universities studied, which corresponds to an effect size d of 0.7 (medium to large) (Figure 7). For smaller effects, there is a higher risk of Type I error.

For detecting an effect size of lower magnitude, the sample size used yields statistical power lower than 80% (60% for an effect size considered as medium, and less than

50% for an effect size considered as small), which is lower than the commonly used threshold.

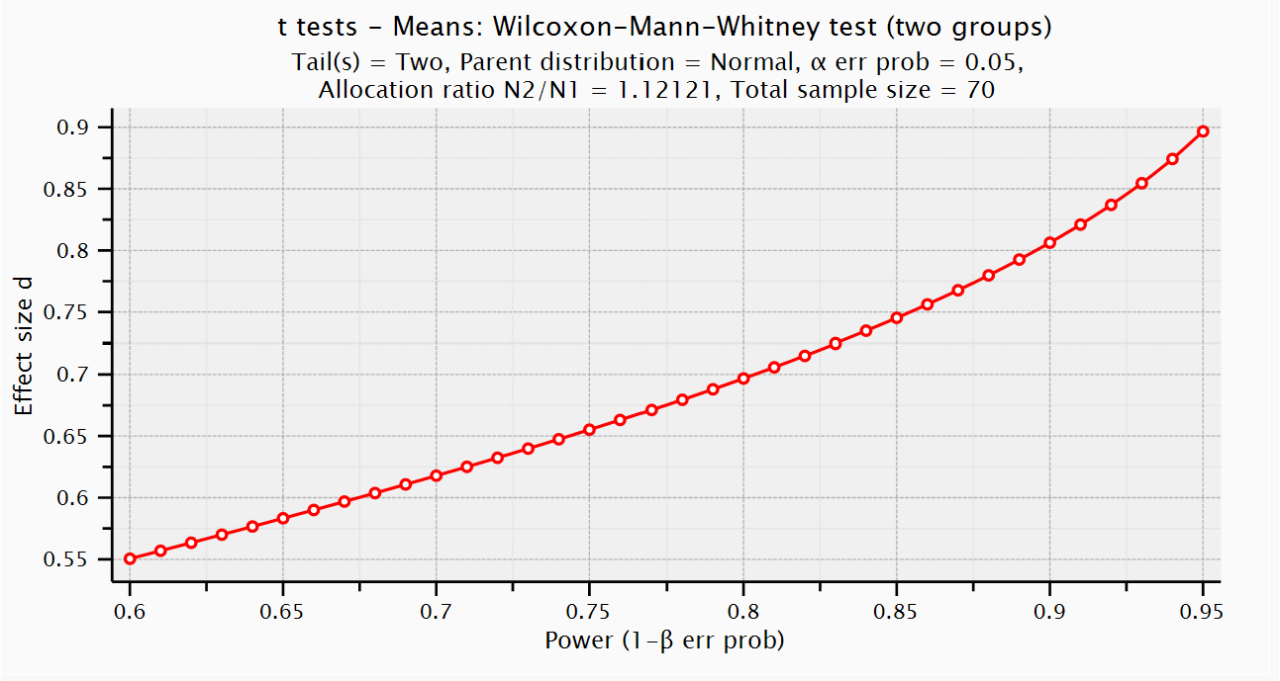


Figure 7 – Sensitivity Analysis for Wilcoxon–Mann–Whitney Test (Two Groups) Using GPower Software.

This graph depicts the relationship between statistical power (1-β error probability) and the minimum detectable effect size (Cohen's d) for a Wilcoxon–Mann–Whitney test with two groups defined as N1=37 and N2=33, corresponding to the simple sizes of students performing OSCE at the University of Turin and University of Parma cohort, respectively. The sensitivity analysis was conducted considering a two-tailed analysis with a normal parent distribution, and an α error probability of 0.05. For a 80% power of the study, the effect size d = 0.7 (medium to large).

	Cohort	OSCE	EFF. Q.	SAT. Q.
Group C vs Group S	UNIPR	**cD** 5 (3.5, 5) vs. 3(3, 3) [0.004]	N.S.	N.S.
Group C vs Group S	UNITO	N.S.	N.S.	**i1** 4(3, 4) vs. 3 (2, 3) [0.005]
Group C + Group S	UNITO vs UNIPR	**cB** 0.19 (0.01, 0.36) [0.034] **cC** 0.40 (0.21, 0.60) [<0.0001]	N.S.	N.S.

Table 13 – Summary of statistically significant differences observed between groups and/or cohorts. Data are presented as median (25th, 75th percentile) and [p value] for comparison between Groups, as β coefficient (95% C.I.) and [p value] for comparison between cohorts. UNIPR and UNITO indicate the University of Parma and Turin, respectively.

Cohort	SAT. Q. X EFF. Q.	SAT. Q. X OSCE	EFF. Q. X OSCE
UNIPR	p = 0.0003 r = 0.74	N.S.	N.S.
UNITO	p < 0.0001 r = 0.75	N.S.	N.S.

Table 14 - Summary of statistically significant association observed within cohorts. Data are obtained with Spearman test. UNIPR and UNITO indicate the University of Parma and Turin, respectively.

DISCUSSION

The 5-point Likert scale was chosen for investigating the perceived efficacy of the proposed practical trainings and the satisfaction of students, due its ability to capture nuanced responses. Moreover, it results easy to use and does not overwhelm respondents with too many choices, thereby improving response rates and increasing the likelihood of genuine answers. The presence of a central option allows participants to express neutrality, avoiding forcing them into agreement or disagreement. However, some limitations are commonly noted in the literature for this type of survey design. First, due to central tendency bias, respondents tend to avoid selecting answers that appear too extreme (e.g., "strongly agree/disagree") or may opt for neutral responses if they are uncertain. Additionally, due to the so-called acquiescence bias, respondents may tend to agree with the given statements regardless of their true opinion, in an attempt to demonstrate a positive disposition toward the questions or to conform to what they believe is expected of them. [37–40]

For the OSCE scoring, a binary system (pass/fail) was chosen for criteria A, B, and C, while a 5-point scoring system (where 1 was “poor” and 5 was “excellent”) was chosen for criterion D. The advantages of the binary system lie in its simplicity and the speed it affords examiners in rating student performance within a restricted available timeframe. Additionally, this scoring method is recognized for its objectivity and for producing lower variability among examiners, which supports scoring consistency. However, the binary scoring system does not capture the varying levels of proficiency in a skill and may not fully represent complex clinical competencies, where performance quality can vary significantly. Furthermore, given the objective to provide students with feedback on their performance, the binary score lacks nuance in conveying how well or poorly a skill was performed. [41,42]

In contrast, the 5-point scoring system allows examiners to capture a range of proficiency, providing a more nuanced evaluation that distinguishes between different levels of competence. Moreover, it facilitates tailored feedback and can more accurately reflect the continuum of clinical competency. However, this system is more

time-consuming, potentially subjective (as examiners may have different interpretations of scale values), and may suffer from “central tendency bias,” causing examiners to avoid extreme scores and cluster ratings around the middle of the scale. From these considerations, it is evident that the choice of the scoring system depends directly on the OSCE context and the evaluation objectives. Therefore, this thesis proposes a hybrid system: the binary scoring system was selected for criteria A, B, and C, which aimed at assessing basic competencies, while a 5-point scale was chosen for criterion D, which aimed at capturing gradations in overall performance. Of particular importance when using a continuous scale is examiner training and calibration, to enhance reliability and validity. In this study, this potential bias was addressed by using the same examiner for each evaluation, along all the study. [43–48]

In consideration of the participation rate in the OSCE at the end of the simulation sessions (Groups S), the noteworthy difference between the two universities could be attributed to the differing logistical organization of the sessions. At the University of Parma, the anatomy dissection laboratory had a lower capacity, which may have led some students to abandon the activity while waiting outside the laboratory for their turn to be evaluated after practicing on the simulator in the Skills Lab. Conversely, at the University of Turin, all students were accommodated together in a larger laboratory and were only allowed to leave the activity once they had completed the OSCE.

Similarly, the difference in the OSCE participation rate between the two groups at the University of Parma can be attributed to the logistical need for Group S to move to a different laboratory between the simulator training activity and the OSCE, as well as the distance between the laboratories, which were located in different buildings in Parma campus, whereas in Turin, they were situated very close to each other. This factor may have led to the loss of some students from the Parma cohort.

One of the objectives of the study was to evaluate whether the different training methods (Group C: cadaveric training, and Group S: simulator training) had a measurable effect on the students’ proficiency in performing nerve blocks on a horse limb, using the OSCE as the method. Another objective was to explore students’

perceived efficacy and satisfaction with the proposed training methods. This was achieved through two tailored questionnaires that students were invited to complete online after the activities.

Once data were collected from the training session in the University of Parma and the University of Turin, the data were analysed to identify any statistically significant differences between groups and cohorts, both in OSCE scores and in questionnaires results.

Regarding the performance results of students from the University of Parma, both groups performed similarly across all OSCE-assessed items, with only one item showing a statistically significant difference. This item was general proficiency in manual skills and techniques, where the cadaver-trained group performed better. This difference can be attributed to the superior haptic feedback provided by cadaveric limbs compared to the simulator. Conversely, the comparable performance between the two groups overall may reflect the strong educational value of simulator training, which effectively simplifies anatomical concepts and allows students to internalize the required manual skills more quickly.

Considering the same comparison from the University of Turin, results showed similar performances across all items assessed in the OSCE for the groups. Those results are summarised in Table 13.

A statistically significant difference was found in the OSCE outcomes of students when comparing the two universities, with more positive results for the students from the University of Turin. This difference can be explained not only by the intrinsic differences between the two populations of students but also by the differences in the didactic curricula at the two universities and the prior experience gained by the students in procedures similar to the proposed activity.

That considered, the training methods seem leading to comparable levels of competency, suggesting that simulator training may be as effective as cadaveric training for developing practical nerve block skills in a controlled setting, and may

provide an adequate or even preferable training option, especially where cadaveric resources are limited, or ethical considerations are a concern.

Another objective of the study was to assess both the satisfaction and perceived efficacy of each training method through post-training questionnaires. Focusing on perceived efficacy, the questionnaire results indicate no difference between the groups at the University of Parma, with all the responses falling between “agree” and “strongly agree” for the proposed statements, indicating a positive overall trend of perceived efficacy. The same outcome was recorded at the University of Turin, with all the responses marked as “agree”.

Regarding the satisfaction questionnaire, no difference between the groups was recorded for the University of Parma cohort, with an overall trend of responses between “agree” and “strongly agree”, indicating a high level of satisfaction with both training methods. The results from the University of Turin were similar, with a similar overall trend; however, a statistically significant difference was found for one question: the Cadaver Group expressed significant higher satisfaction with the effectiveness of the training in achieving the learning objectives compared to the Simulator Group. This difference may be attributed to the low fidelity of the simulator in providing haptic feedback during nerve palpation. This consideration was also confirmed by the free comment received in the questionnaires. Nevertheless, the overall high satisfaction expressed by the participants in both groups underscores the effectiveness of both training methods in achieving positive educational outcomes. Those results are summarised in Table 13.

A positive correlation was found between the results of the Satisfaction and Efficacy questionnaires for both the Universities cohort, indicating that the students who appreciated the activity the most, were also the ones who felt they learned the most. Results are summarised in Table 14.

An aspect to consider is the different response rates to the questionnaires between the two universities. First, it should be noted that filling out the questionnaires was not compulsory for the students, as it was not part of their academic curricula, although

they were sensitised to the importance of their collaboration for the experiment, both at the end of the lecture and during the practical activity. Moreover, in contrast to other similar experiments, no incentives to participate were offered to the participants. The lectures, practical sessions and OSCEs were conducted by the same instructors and professor for both cohorts. Despite this, it can also be hypothesized that the two student populations differ in terms of their interest in the subject and their involvement in collaborating on the project.

In light of the results obtained from the sensitivity analyses conducted a posteriori on the assessed population, which showed sufficient power of our study only for detecting a “very large” effect size (Cohen’s d) for comparisons within groups of both cohorts, and a “large” effect size for comparisons between cohorts, it is important to note that the observed results may be influenced by the limited sample size. This finding suggests the need to expand the student sample size in future work, to address this limitation.

Another limitation of the study to consider is the execution of the OSCE on cadaveric limbs for both groups: the students trained on cadavers and those trained on simulators. This approach, chosen to standardize the outcome by measuring the performance of both groups using the same method and in the most realistic scenario possible—given that, for ethical reasons, the task under study cannot be performed on live animals—could have introduced a bias. Specifically, only one group of students had prior exposure to the evaluative scenario during training. Considering this, the statistically significant difference observed at the University of Parma in the OSCE item related to overall general proficiency could be attributed, on one hand, to the simulator's low fidelity in replicating real anatomy, and on the other hand, to the previously described bias. However, this result was not confirmed in the cohort of students at the University of Turin. At the same time, the comparable results observed for the other items in the Parma cohort, as well as for all items in the Turin cohort, could indicate a minimal effect of the described bias. An additional prospective re-assessment could explore the impact of the two different training methods on long-term skill retention and real-world

clinical outcomes. To this aim, the recent introduction of the *Tirocinio pratico valutativo* (an evaluative practical internship) in the 5th-year course could provide relevant data. The above-mentioned internship aims to develop and assess professional skills and competencies across various sectors of veterinary practice, in alignment with the "Day One Competences" established by the European Association of Establishments for Veterinary Education (EAEVE)[8] and European regulations [49].

[8] EAEVE / AEEEV © 2024 - https://www.eaeve.org/

CONCLUSION

This study provides substantial evidence supporting the efficacy of simulation-based medical education (SBME) in veterinary training, particularly for developing core competencies required for equine practice. Through structured comparisons between cadaveric and simulator-based training for performing equine distal limb nerve blocks, this research underscores the viability of simulators as a complementary or even alternative training tool to cadaveric limbs, especially in educational contexts where ethical, logistical, or resource constraints limit cadavers' availability.

The results showed that both cadaveric and simulator training approaches lead to comparable proficiency in nerve block techniques, as evidenced by Objective Structured Clinical Examination (OSCE) outcomes across multiple institutions. While cadaver-based training slightly enhanced students' skill proficiency due to superior haptic feedback, the simulator effectively simplified anatomical learning, enabling students to effectively internalize concepts. This validates the role of simulators in teaching Day One Competencies, contributing to the fundamental skills in equine practice.

However, the marginally lower satisfaction ratings regarding tactile realism in simulator training suggest the need for improvement through the development of a more high-fidelity model that more accurately replicates the tactile feedback of real anatomy. Such advancements could enable students to practice core skills autonomously, ensuring they meet curriculum objectives and demonstrate readiness for unsupervised clinical tasks, particularly in the evaluative phase of their education (*tirocinio pratico valutativo*). Enhancing simulator fidelity would thus support not only initial skill acquisition but also long-term skill retention and professional competency, aligning with European veterinary education standards and the Day One Competencies framework.

Moreover, this work aligns with the European Commission's ethical commitment to the "3R" principles (Replacement, Reduction, and Refinement), advocating for

humane and effective alternatives in veterinary training [30]. By offering a standardized, accessible, and ethically grounded approach, the study contributes to a broader effort to reduce the dependency on cadaveric specimens, supporting sustainable and ethical practices within veterinary education.

Future research should focus on increasing the sample size to substantiate these findings further, as well as on investigating the long-term impacts of simulator-based training on clinical performance. A prospective study evaluating skill retention and real-world application would provide valuable insights into the durability of simulator-acquired skills. In this way, ongoing advancements in SBME can continue to enhance veterinary education, meeting the evolving demands of both the veterinary field and ethical standards.

New horse distal limb simulator for nerve blocks simulation and ultrasound training on flexor metacarpal region

As a collateral and additional project to this thesis, the author is collaborating on the development of a high-fidelity equine distal limb simulator capable of simulating the nerve block procedure and, additionally, training students in the ultrasonography of the flexor metacarpal region of the horse. The objective of this project, which is still under development, is to create a simulator that can overcame the limitations of currently available commercial products.

For the development of that project, a high-definition CT scan of a cadaveric front leg was obtained in collaboration with the University of Parma.

The scan was segmented to create virtual 3D models of the bones in the limb up to the metacarpus (including the first, second, and third phalanges, navicular bone, sesamoid bones, and principal and accessory metacarpal bones). The Superficial Digital Flexor Tendon (SDFT), Deep Digital Flexor Tendon (DDFT), and Suspensory Ligament (SL) were also segmented.

After a deep analysis of the literature and research on the available materials and technologies for meeting the project objectives, 3DZ Group[9], a leading additive manufacturing company in Europe, was selected as a partner for the simulator development.

In collaboration with 3DZ engineers, additive manufacturing (3D printing) and injection molding were identified as suitable technologies for creating the simulator.

Various materials were tested for printing the simulator components, especially with respect to their ultrasound appearance. For the bones and hoof, PLA was selected due to its durability. Flexible resin (Formlabs Flexible 80A and Elastic 50A) was chosen to print tendons and ligaments because of its elastomeric proprieties.

Tests for echogenicity showed that PLA completely reflects ultrasound waves, while the flexible resin provided a texture and a greyscale similar to that of tendons and ligaments in ultrasound examinations.

Additionally, ballistic gel was identified as a cost-effective material that is permeable to ultrasound waves, inert, stable at room temperature, and suitable for representing the remaining soft tissues.

For the external finish, platinum silicone, often used in cinematographic effects, was chosen for its high level of realism.

Once the components will be ready, they will be integrated with a circuit simulating the target nerves, capable of giving both an auditory and visual (LED) feedback when the needle is correctly inserted.

The final assembly will be mounted inside a bivalve 3D-printed mold, into which molten ballistic gel will be injected in a vacuum chamber to remove trapped air and allow for a clear ultrasound imaging.

Finally, the simulator will be coated with platinum silicone to provide a realistic appearance and feel.

[9] 3DZ spa - Via dei Pini 32, Treviso · ITALY

Once the first prototype will be completed, it will be tested by students under the same condition as the commercial simulator. The results from the cadaveric leg training, the commercial simulator and the new prototype will be compared. The prototype will also be evaluated by a pool of senior veterinary experts.

BIBLIOGRAPHY

1. Noyes, J.A., Carbonneau, K.J. and Matthew, S.M. (2022) Comparative Effectiveness of Training with Simulators Versus Traditional Instruction in Veterinary Education: Meta-Analysis and Systematic Review. *J Vet Med Educ* **49**, 25–38.

2. Ker, J. and Bradley, P. (2013) Simulation in medical education. *Understanding Medical Education: Evidence, Theory and Practice: Second Edition* 175–192. https://onlinelibrary.wiley.com/doi/full/10.1002/9781118472361.ch13. Accessed August 27, 2024.

3. Gordon, S., Gardner, D., Weston, J., Bolwell, C., Benschop, J. and Parkinson, T. (2021) Fostering the Development of Professionalism in Veterinary Students: Challenges and Implications for Veterinary Professionalism Curricula. *Education Sciences 2021, Vol. 11, Page 720* **11**, 720. https://www.mdpi.com/2227-7102/11/11/720/htm. Accessed October 11, 2024.

4. European Association of Establishments for Veterinary Education EAEVE / AEEEV. https://www.eaeve.org/. Accessed January 17, 2025.

5. AAVMC, Hunt, J., Delcambre, J. and Baillie, S. (2024) AAVMC Use of Animals in Veterinary Education Handbook. https://osf.io/5k894. Accessed January 17, 2025.

6. AAVMC Task Force on the Use of Animals in Veterinary Education, Hunt, J.A., Baillie, S., Beaver, B.V., Carey, A.M.J., Cary, J.A., Delcambre, J., Foster, D.M., Hendrickson, D.A., Huxley, J.N. and Mashima, T.Y. (2022) Guidelines for the Use of Animals in Veterinary Education. *American Association of Veterinary Medical Colleges.* . https://www.aavmc.org/wp-content/uploads/2023/05/AAVMC-Guidelines-AnimalUse.pdf. Accessed January 17, 2025.

7. Andersen, A.G., Rahmoui, L., Dalsgaard, T.S., Svendsen, M.B.S., Clementsen, P.F., Konge, L. and Bjerrum, F. (2023) Preparing for Reality: A Randomized Trial

on Immersive Virtual Reality for Bronchoscopy Training. *Respiration* **102**, 316–323. https://pubmed.ncbi.nlm.nih.gov/36796339/. Accessed October 11, 2024.

8. Zemljic, B. (2004) Standards of Veterinary Education for the Future: Preparing the Profession for the New Century. *https://doi.org/10.3138/jvme.31.1.13* **31**, 13–14. https://utppublishing.com/doi/10.3138/jvme.31.1.13. Accessed January 15, 2025.

9. Braid, H.R. (2022) The Use of Simulators for Teaching Practical Clinical Skills to Veterinary Students - A Review. *Altern Lab Anim* **50**, 184–194.

10. Elshama, S.S. and Elshama, S.S. (2020) How to apply Simulation-Based Learning in Medical Education? *Iberoamerican Journal of Medicine* **2**, 79–86. http://www.iberoamjmed.com/article/doi/10.5281/zenodo.3685233. Accessed May 15, 2024.

11. Parasuraman, R., Tippelt, R. and Hellwig, L. (2007) Brain, Cognition and Learning in Adulthood. *Understanding the Brain: The Birth of a Learning Science* 211–237.

12. Wetzel, C.M., Kneebone, R.L., Woloshynowych, M., Nestel, D., Moorthy, K., Kidd, J. and Darzi, A. (2006) The effects of stress on surgical performance. *Am J Surg* **191**, 5–10. https://pubmed.ncbi.nlm.nih.gov/16399098/. Accessed September 9, 2024.

13. Williamson, J.A., Dascanio, J.J., Christmann, U., Johnson, J.W., Rohleder, B. and Titus, L. (2016) Development and Validation of a Model for Training Equine Phlebotomy and Intramuscular Injection Skills. *J Vet Med Educ* **43**, 235–242. https://pubmed.ncbi.nlm.nih.gov/27075276/. Accessed October 31, 2024.

14. Fox, V., Sinclair, C., Bolt, D.M., Lowe, J. and Weller, R. (2013) Design and validation of a simulator for equine joint injections. *J Vet Med Educ* **40**, 152–157. https://pubmed.ncbi.nlm.nih.gov/23709111/. Accessed October 31, 2024.

15. Gunning, P., Smith, A., Fox, V., Bolt, D.M., Lowe, J., Sinclair, C., Witte, T.H. and Weller, R. (2013) Development and validation of an equine nerve block simulator

to supplement practical skills training in undergraduate veterinary students. *Vet Rec* **172**, 450. https://pubmed.ncbi.nlm.nih.gov/23486507/. Accessed October 31, 2024.

16. Elnady, F.A., Sheta, E., Khalifa, A.K. and Rizk, H. (2015) Training of upper respiratory endoscopy in the horse using preserved head and neck. *ALTEX* **32**, 384–387. https://pubmed.ncbi.nlm.nih.gov/26214852/. Accessed October 31, 2024.

17. Nagel, C., Ille, N., Erber, R., Aurich, C. and Aurich, J. (2015) Stress Response of Veterinary Students to Gynaecological Examination of Horse Mares - Effects of Simulator-Based and Animal-Based Training. *Reprod Domest Anim* **50**, 866–871. https://pubmed.ncbi.nlm.nih.gov/26310966/. Accessed October 31, 2024.

18. Nagel, C., Ille, N., Aurich, J. and Aurich, C. (2015) Teaching of diagnostic skills in equine gynecology: simulator-based training versus schooling on live horses. *Theriogenology* **84**, 1088–1095. https://pubmed.ncbi.nlm.nih.gov/26174037/. Accessed October 31, 2024.

19. Elarbi, M.M., Ragle, C.A., Fransson, B.A. and Farnsworth, K.D. (2018) Face, construct, and concurrent validity of a simulation model for laparoscopic ovariectomy in standing horses. *J Am Vet Med Assoc* **253**, 92–100. https://pubmed.ncbi.nlm.nih.gov/29911940/. Accessed October 31, 2024.

20. Allavena, R.E., Schaffer-White, A.B., Long, H. and Alawneh, J.I. (2017) Technical Skills Training for Veterinary Students: A Comparison of Simulators and Video for Teaching Standardized Cardiac Dissection. *J Vet Med Educ* **44**, 620–631. https://pubmed.ncbi.nlm.nih.gov/28581914/. Accessed October 31, 2024.

21. Wass, V., Van Der Vleuten, C., Shatzer, J. and Jones, R. (2001) Assessment of clinical competence. *Lancet* **357**, 945–949.

22. Berendes, S., Schaper, E., Tipold, A. and Wissing, S. (2023) Evaluation of the eOSCE for testing clinical skills. *Front Vet Sci* **10**, 1196311.

23. Hunt, J.A., Simons, M.C. and Anderson, S.L. (2022) If you build it, they will learn: A review of models in veterinary surgical education. *Veterinary Surgery* **51**, 52–61. https://onlinelibrary.wiley.com/doi/full/10.1111/vsu.13683. Accessed September 9, 2024.

24. MOTION FOR A RESOLUTION on plans and actions to accelerate the transition to innovation without the use of animals in research, regulatory testing and education | B9-0428/2021 | European Parliament . https://www.europarl.europa.eu/doceo/document/B-9-2021-0428_EN.html. Accessed October 10, 2024.

25. The Constitution of the Italian Republic | www.governo.it . https://www.governo.it/en/fundamental-principles/11825. Accessed October 10, 2024.

26. Zemanova, M.A., Knight, A. and Lybæk, S. (2021) Educational use of animals in Europe indicates a reluctance to implement alternatives. *ALTEX* **38**, 490–506. https://pubmed.ncbi.nlm.nih.gov/33837436/. Accessed October 10, 2024.

27. McCool, K.E., Bissett, S.A., Hill, T.L., Degernes, L.A. and Hawkins, E.C. (2020) Evaluation of a Human Virtual-Reality Endoscopy Trainer for Teaching Early Endoscopy Skills to Veterinarians. *J Vet Med Educ* **47**, 106–116. https://pubmed.ncbi.nlm.nih.gov/31009293/. Accessed October 11, 2024.

28. McCool, K.E., Marks, S.L. and Hawkins, E.C. (2022) Endoscopy Training in Small Animal Internal Medicine: A Survey of Residency Training Programs in North America. *J Vet Med Educ* **49**, 515–523. https://pubmed.ncbi.nlm.nih.gov/34115581/. Accessed October 11, 2024.

29. European Parliament (2021) Resolution on plans and actions to accelerate the transition to innovation without the use of animals in research, regulatory testing and education - 2021/2784(RSP). https://oeil.secure.europarl.europa.eu/oeil/en/procedure-file?reference=2021/2784(RSP). Accessed January 18, 2025.

30. Maestri, E. (2021) The 3Rs Principle in Animal Experimentation: A Legal Review of the State of the Art in Europe and the Case in Italy. *BioTech* **10**, 9. https://pmc.ncbi.nlm.nih.gov/articles/PMC9245475/. Accessed October 30, 2024.

31. Starke, S.D. and May, S.A. (2017) Veterinary student competence in equine lameness recognition and assessment: a mixed methods study. *Vet Rec* **181**. https://pubmed.ncbi.nlm.nih.gov/28801497/. Accessed October 12, 2024.

32. University of Georgia (2019) *Diagnostic Analgesia of the Equine Distal Limb*, Apple Books. https://books.apple.com/it/book/diagnostic-analgesia-of-the-equine-distal-limb/id1449607873. Accessed October 20, 2024.

33. Life sized 3D replicas of horse foot anatomy - 3D Vet Anatomize . https://3dvetanatomize.com/. Accessed October 20, 2024.

34. Rajaure, Y.S., Thapa, B., Budhathoki, L., Rana, S.R.L., Neupane, R. and Karki, P. (2023) Assessment of performance and confidence level of simulation based clinical examination of respiratory system in undergraduate medical students of a medical college: A comparative cross-sectional study. *Annals of Medicine & Surgery* **86**, 756–760. https://journals.lww.com/10.1097/MS9.0000000000001631.

35. Artino, A.R., La Rochelle, J.S., Dezee, K.J. and Gehlbach, H. (2014) Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach* **36**, 463. /pmc/articles/PMC4059192/. Accessed August 14, 2024.

36. Zach Bobbitt (2020) Effect Size: What It Is and Why It Matters. https://www.statology.org/effect-size/. Accessed October 25, 2024.

37. DeVellis, R.F.. (2016) Scale development : theory and applications. 280. https://books.google.com/books/about/Scale_Development.html?hl=it&id=48ACCwAAQBAJ. Accessed October 23, 2024.

38. Likert, R. (1932) A technique for the measurement of attitudes. *Archives of psychology* .

39. Joshi, A., Kale, S., Chandel, S. and Pal, D.K. (2015) Likert Scale: Explored and Explained. *Current Journal of Applied Science and Technology* **7**, 396–403. https://journalcjast.com/index.php/CJAST/article/view/381. Accessed October 23, 2024.

40. Krosnick, J.A. and Presser, S. (2010) Question and Questionnaire Design. In: *Handbook of Survey Research*, Second., Emerald Group Publishing Limited.

41. Chahine, S., Holmes, B. and Kowalewski, Z. (2016) In the minds of OSCE examiners: uncovering hidden assumptions. *Advances in Health Sciences Education* **21**, 609–625. https://link.springer.com/article/10.1007/s10459-015-9655-4. Accessed October 30, 2024.

42. Piumatti, G., Cerutti, B. and Perron, N.J. (2021) Assessing communication skills during OSCE: need for integrated psychometric approaches. *BMC Med Educ* **21**, 1–10. https://link.springer.com/articles/10.1186/s12909-021-02552-8. Accessed October 30, 2024.

43. Pell, G., Fuller, R., Homer, M. and Roberts, T. (2010) How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. *Med Teach* **32**, 802–811. https://pubmed.ncbi.nlm.nih.gov/20854155/. Accessed October 27, 2024.

44. Ronald M. Harden, Pat Lilley, M.P. (2015) *The Definitive Guide to the OSCE. The Objective Structured Clinical Examination as a performance assessment*, Elsevier.

45. Ferris, H.A. and Flynn, D.O. (2015) Assessment in Medical Education; What Are We Trying to Achieve? *International Journal of Higher Education* **4**, 139. https://www.sciedu.ca/journal/index.php/ijhe/article/view/6662. Accessed October 27, 2024.

46. NEWBLE, D.I. and SWANSON, D.B. (1988) Psychometric characteristics of the objective structured clinical examination. *Med Educ* **22**, 325–334. https://pubmed.ncbi.nlm.nih.gov/3173161/. Accessed October 27, 2024.

47. Norcini, J.J. and McKinley, D.W. (2007) Assessment methods in medical education. *Teach Teach Educ* **23**, 239–250.

48. Swanson, D.B., Norman, G.R. and Linn, R.L. (1995) Performance-Based Assessment: Lessons From the Health Professions. *http://dx.doi.org/10.3102/0013189X024005005* **24**, 5–11. https://journals.sagepub.com/doi/abs/10.3102/0013189X024005005. Accessed October 27, 2024.

49. May, S.A. and Head, S.D. (2010) Assessment of technical skills: Best practices. *J Vet Med Educ* **37**, 258–265. https://jvme.utpjournals.press/doi/10.3138/jvme.37.3.258. Accessed July 3, 2024.