education sciences

Article

Learner Satisfaction with Technical and Non-Technical Skills in an Accredited Healthcare Simulation Centre: A Retrospective Study

Silvia Reverté-Villarroya [1,2,3], Gerard Mora-Lopez [1,2,*], Laia Grau-Castell [4,*], Jose Maria Esteban Blanco [3], Jordi Baucells Rodriguez [3], Ramon Palau [4,5] and Mireia Adell-Lleixà [1,2]

1 Nursing Department, Faculty of Nursing, Universitat Rovira i Virgili, Campus Terres de l'Ebre, 43500 Tortosa, Spain; silvia.reverte@urv.cat (S.R.-V.); mireia.adell@urv.cat (M.A.-L.)
2 Research Group on Advanced Nursing (CARING)-161, Universitat Rovira i Virgili, 43002 Tarragona, Spain
3 Càtedra d'Innovació i Simulació en Salut (CISTE), Fundació URV, Campus Terres de l'Ebre, Avenue Remolins 13-15, 43500 Tortosa, Spain; josemaria.esteban@fundacio.urv.cat (J.M.E.B.); jordi.baucells@fundacio.urv.cat (J.B.R.)
4 Department of Pedagogy, Universitat Rovira i Virgili, Campus Terres de l'Ebre, 43500 Tortosa, Spain; ramon.palau@urv.cat
5 ARGET Research Group, Faculty of Education Sciences and Psychology, Universitat Rovira i Virgili, 43007 Tarragona, Spain
* Correspondence: gerard.mora@urv.cat (G.M.-L.); laia.grauc@urv.cat (L.G.-C.)

Abstract

Clinical simulation is a cornerstone of healthcare education, yet comparative evidence on learner perceptions of technical skills (TS) versus non-technical skills (NTS) training remains limited. This study aimed to examine Kirkpatrick Level 1 satisfaction across simulation course types and explore associations with participant characteristics. We conducted a retrospective analysis of 1806 healthcare professionals who attended 133 simulation sessions between 2024 and 2025. Participants completed a 16-item ad hoc questionnaire (Cronbach $\alpha = 0.959$) covering four domains scored 1–10. Domain scores were compared between TS ($n = 1224$) and NTS ($n = 582$) courses using Welch t-tests and Cohen's d. Spearman correlations assessed associations with age, gender, and professional category. Overall learner satisfaction was excellent (mean 9.55 ± 0.70). TS courses scored significantly higher across all domains ($p < 0.001$), with effect sizes ranging from small (d = 0.21–0.28) to moderate-approaching (d = 0.45) in the Simulation/Methodology domain. Item-level analysis highlighted perceived improvement in technical skills (d = 0.49) and error awareness (d = 0.43) as the most discriminating factors. Age, gender, and professional category were significantly associated with course type (all $p \leq 0.020$). Both TS and NTS simulations achieve high learner satisfaction. The perceptual advantage of TS courses suggests that instructional strategies enhancing the tangibility of NTS learning outcomes are warranted to optimize learner engagement and satisfaction.

Keywords: clinical simulation; professional competence; health personnel attitudes

Check for updates

Academic Editor: Andrew Williams

Received: 22 April 2026
Revised: 13 May 2026
Accepted: 19 May 2026
Published: 21 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Clinical simulation has become a cornerstone of healthcare education over the past two decades, providing immersive, controlled environments in which learners can practise clinical tasks and decision-making without risk to patients. Simulation-based education (SBE) has been shown to enhance technical proficiency, reduce medical errors, and contribute

to improved patient safety across healthcare disciplines (Muñoz Gualán & Sierra, 2025). Evidence from systematic reviews confirms that SBE significantly improves both technical competencies and a range of non-technical skills including communication, teamwork, leadership, and clinical reasoning (Muñoz Gualán & Sierra, 2025; Patten et al., 2026; Pucer et al., 2025).

Despite this strong evidence base, a persistent asymmetry exists in the simulation literature: the majority of research and curricular investment has focused on technical skills (TS) training, while non-technical skills (NTS), encompassing interpersonal, cognitive, and behavioural competencies such as situational awareness, decision-making, and crisis resource management, remain comparatively understudied (Patten et al., 2026; Pucer et al., 2025). This imbalance is particularly concerning given that adverse clinical events are frequently attributed to failures in NTS rather than technical errors (Gawronski et al., 2022). Some authors argue that NTS are best developed through interprofessional simulation, yet others highlight the lack of validated measurement instruments and the methodological heterogeneity across studies as barriers to drawing firm conclusions (Patten et al., 2026; Gawronski et al., 2022). It is worth noting that several behavioural assessment tools for NTS have been developed and validated in specific clinical contexts, including Non-Technical Skills for Crew Resource Management (NOTECHS) (Flin et al., 2003), Anaesthetists' Non-Technical Skills (ANTS) (Fletcher et al., 2003), and the Ottawa Crisis Resource Management Global Rating Scale (Ottawa GRS) (Kim et al., 2006; Kim et al., 2009). These instruments provide structured, observer-rated evaluations of NTS competencies during simulation, in contrast to the learner self-report perception measures used in the present study. Their existence contextualises the current methodological choice: while validated behavioural rating tools capture objective performance, perception-based measures such as those employed here capture a distinct and complementary construct, the learner's subjective experience of training quality, which is the appropriate operationalisation of Kirkpatrick Level 1 evaluation.

A key but underexplored question is whether learners perceive TS and NTS simulation training as equally effective. Understanding learner satisfaction is directly relevant to educational practice because, within Kirkpatrick's four-level evaluation framework (Kirkpatrick & Kirkpatrick, 2016), Level 1 reactions are considered foundational prerequisites for deeper learning and skill transfer. Positive learner perceptions have been linked to increased engagement and more effective application of skills in clinical practice (Bdiri Gabbouj et al., 2024; McGaghie et al., 2010). However, whether participants in TS and NTS courses differ systematically in their Level 1 reactions, and whether these differences have practical educational significance beyond statistical significance, has not been rigorously examined.

Furthermore, demographic and professional characteristics may moderate how learners engage with and perceive different types of simulation training. Age, professional category, and institutional setting have been shown to shape educational preferences and participation patterns in simulation programmes (Zhang, 2023; Pimenta et al., 2025; Prete et al., 2024; Ylönen et al., 2025), yet comprehensive analyses of these moderating effects across TS and NTS course types remain scarce, especially in multi-institutional, accredited simulation settings. Clarifying these associations has direct implications for equitable programme design and targeted learner recruitment.

Addressing these gaps, the present study aimed to: (1) analyse healthcare professionals' perceptions of simulation training classified by primary learning objective (TS versus NTS) using Kirkpatrick's Level 1 framework; (2) examine the educational significance of perception differences by calculating Cohen's d effect sizes at both domain and item level; and (3) explore associations between participant characteristics and course type at an SSH-accredited clinical simulation centre serving multiple healthcare institutions in Spain.

2. Materials and Methods

2.1. Study Design, Setting, and Ethics

This retrospective cross-sectional study included all simulation-based training activities conducted at the Centre d'Innovació i Simulació Territorial (CISTE) during 2024 and 2025. CISTE is a high-performance clinical simulation centre affiliated with the Chair of Innovation and Simulation in Health of les Terres de l'Ebre (Universitat Rovira i Virgili, Spain), accredited by the Society for Simulation in Healthcare (SSH) and operating in accordance with SSH and INACSL Standards of Best Practice in all training activities. The retrospective design was chosen to maximise the available sample from routinely collected institutional data without introducing selection effects that a prospective design might create.

Ethical approval was granted by the Research Ethics Committee on Medicinal Products (Comitè d'Ètica d'Investigació amb Productes Sanitaris i de la Salut d'Atenció Primària, CEIPSA) under project code 2023 TDO 0112. All participants provided written informed consent prior to inclusion. The study was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013) and applicable Spanish data protection regulations (Organic Law 3/2018).

2.2. Course Classification and Participants

All healthcare professionals and students enrolled in simulation training at CISTE during the study period were eligible for inclusion. Courses were classified according to their primary learning objective as either technical skills (TS) or non-technical skills (NTS) training. TS courses focused on procedural competencies and hands-on clinical technique acquisition using high-fidelity mannequins and task trainers (e.g., advanced life support, surgical skills, ultrasound-guided procedures). NTS courses targeted cognitive, behavioural, and interpersonal competencies, including communication, teamwork, leadership, and crisis resource management.

For the 2025 dataset (n = 941), course classification was assigned prospectively by the CISTE expert team and recorded in the institutional database at the time of programme design. For the 2024 dataset (n = 865), classification was applied retrospectively by the same multidisciplinary team (comprising two simulation educators and one clinical specialist) using course names, syllabi, and stated learning objectives. Consensus was reached for all 45 courses; no course required arbitration or was excluded due to ambiguity. The classification criterion was identical across both years: a course was assigned to TS if its primary learning objective was procedural competency acquisition, and to NTS if its primary objective was the development or application of cognitive, interpersonal, or behavioural competencies. No course was classified as both.

Participant demographic and professional data collected at enrolment included age (date of birth), biological sex, professional category (nurse, physician, TCAI, student, or other), and institution type (hospital, intermediate care, mental health, or university/community settings). Age was computed from date of birth and date of the session; values outside the plausible range (17–66 years) were excluded from age-related analyses (n = 1547 valid, 85.7%).

2.3. Data Collection Instrument

Participant perceptions were assessed at the end of each training session using a 16-item ad hoc questionnaire administered electronically via Microsoft Forms (Microsoft Corporation, Redmond, WA, USA). The instrument was structured into four domains: (1) Simulation/Methodology (8 items; e.g., "The simulation session met my learning expectations"; "The simulation improved my awareness of clinical errors"); (2) Teachers/Instructors

(3 items; e.g., "Instructors created a safe learning environment"); (3) Technological Materials (2 items; e.g., "The simulated environment was realistic"); and (4) Facilities/Staff (3 items; e.g., "The facilities met my expectations"). Each item was rated on a 10-point Likert-type scale (1 = strongly disagree; 10 = strongly agree). Completion was voluntary and anonymous.

Because the questionnaire was developed specifically for this study, the following structural validation analyses were conducted to establish its psychometric suitability as the primary measurement instrument of the study: internal consistency, exploratory factor analysis, and convergent and discriminant validity assessment. Internal consistency was assessed using Cronbach's alpha (α) for the total scale and each domain separately. The total scale demonstrated excellent reliability ($\alpha = 0.958$), exceeding the $\alpha > 0.90$ threshold recommended for research instruments (Taber, 2018). Domain-level alpha values were: Simulation/Methodology $\alpha = 0.930$; Teachers/Instructors $\alpha = 0.898$; Technological Materials $\alpha = 0.748$; Facilities/Staff $\alpha = 0.912$. The questionnaire instrument, item-level data, and full EFA results are available in Supplementary Table S1.

2.3.1. Structural Validity: Exploratory Factor Analysis

An EFA was conducted on the complete-case item correlation matrix (N = 1814 participants with all 16 items present). Suitability was confirmed by KMO = 0.957 ("meritorious"; Kaiser, 1974) and Bartlett's test of sphericity ($\chi^2(120) = 26{,}131.5$, $p < 0.001$). The number of factors was determined by parallel analysis (200 random datasets; 95th percentile threshold): only F1 ($\lambda = 10.143$) exceeded its random counterpart ($\lambda = 1.196$), while F2 ($\lambda = 1.127$) fell below the parallel analysis threshold ($\lambda = 1.155$). A one-factor solution was therefore retained. Extraction was performed using Maximum Likelihood (ML). The single factor explained 63.4% of total variance. All 16 items showed substantial loadings on F1 (range: 0.493–0.849; all ≥ 0.40), with mean communality = 0.475. Full factor loadings and communalities are reported in Supplementary Table S1.

These results indicate that the 16-item questionnaire is essentially unidimensional: all items measure a single general satisfaction factor. The overall perception score (mean of all 16 items) is therefore the psychometrically justified primary outcome variable. Domain-level groupings (Simulation/Methodology, Teachers/Instructors, Technological Materials, Facilities/Staff) function as analytical facets of this single construct rather than as independent latent dimensions, and domain-level comparisons should be interpreted accordingly.

2.3.2. Convergent and Discriminant Validity

Convergent validity was assessed using the average variance extracted (AVE) per domain (Fornell & Larcker, 1981). Only the Simulation/Methodology domain met the AVE ≥ 0.50 threshold (AVE = 0.564); the remaining domains showed AVE values below 0.50 (Teachers/Instructors: 0.339; Technological Materials: 0.459; Facilities/Staff: 0.386), reflecting the strong cross-domain correlations (r = 0.71–0.80) characteristic of a unidimensional instrument. Composite reliability (CR) values ranged from 0.602 to 0.911. The Fornell–Larcker discriminant validity criterion was not met for any domain ($\sqrt{\text{AVE}} = 0.582$–$0.751$; all below maximum inter-domain r = 0.803), confirming that the four subscales do not constitute empirically distinct constructs. These findings are consistent with the unidimensional EFA solution and further support using the overall score as the primary outcome. Detailed indices are provided in Supplementary Table S1.

2.4. Statistical Analysis

All analyses were performed using IBM SPSS Statistics version 27.0 (IBM Corporation, Armonk, NY, USA). Analyses were conducted at the individual participant level (n = 1806). Descriptive statistics were calculated for all variables; continuous variables are reported

as mean ± standard deviation (SD) and median; categorical variables as frequencies and percentages.

Associations between course type (TS vs. NTS) and participant characteristics (age, gender, professional category, and institution type) were examined using Spearman's rank correlation coefficients (ρ) and chi-square (χ^2) tests. Where chi-square indicated a significant association but the Spearman correlation did not, the association was interpreted as non-ordinal. Group differences in domain perception scores were assessed using independent-samples t-tests with Welch's correction for unequal variances; a two-sided $p < 0.05$ was considered statistically significant.

To quantify the practical educational significance of observed differences, Cohen's d effect sizes were calculated for all TS vs. NTS comparisons at both domain and item level, using pooled standard deviations. Effect sizes were interpreted as trivial (d < 0.20), small (d = 0.20–0.49), medium (d = 0.50–0.79), or large (d ≥ 0.80) (Cohen, 1988). This step was considered essential given the large sample size, which confers sufficient statistical power to detect educationally trivial differences as statistically significant.

A post hoc sample size adequacy check was performed using G*Power 3.1 (Faul et al., 2007). For the t-test analyses, the minimum required sample was estimated based on the smallest observed effect size (d = 0.21, Facilities/Staff domain), an allocation ratio of 1:2.1 (TS:NTS), $\alpha = 0.05$ (two-tailed), and 80% power, yielding a minimum of 796 participants. The enrolled sample (n = 1806) exceeded this by a factor of 2.3, with observed power ranging from 98.8% to >99.9% across domains and a minimum detectable effect of d = 0.14. For Spearman correlations, the Fisher z-transformation method indicated a minimum requirement of n = 490 for the weakest significant association (ρ = 0.126); the enrolled sample provided adequate power for all significant correlations (ρ = 0.059–0.107).

3. Results

3.1. Sample Characteristics

A total of 1806 participants were included, corresponding to 133 simulation sessions conducted in 2024 ($n = 865$; 45 sessions) and 2025 ($n = 941$; 88 sessions). Technical skills (TS) courses accounted for 1224 participants across 94 sessions, while non-technical skills (NTS) courses included 582 participants across 39 sessions (Table 1). Age data were available for 1547 participants (85.7%). The median age of the sample was 34 years. Participants in TS courses were younger than those in NTS courses (median 32 vs. 36 years; Mann–Whitney U, $p = 0.020$). Overall, 1440 participants (79.7%) were female. The proportion of women was higher in TS courses (81.5%) than in NTS courses (75.9%), and gender was significantly associated with course type ($\chi^2 = 21.795$, $p < 0.001$). Professional category differed significantly between course types ($\chi^2 = 188.003$, $p < 0.001$), with physicians and nurses more represented in TS courses, whereas participants classified as "other" were more frequent in NTS courses. Institution type was also associated with course type ($\chi^2 = 50.568$, $p < 0.001$); institution data were available for 1514 participants (83.8%).

Table 1. Sample characteristics by course type (TS vs. NTS).

Variable	Total (N = 1806)	TS (n = 1224; 94 Sessions)	NTS (n = 582; 39 Sessions)	p-Value
Age, years †	34	32	36	0.020 ‡
Female, n (%)	1.440 (79.7)	997 (81.5)	443 (75.9)	<0.001 §
Professional category, n (%)				<0.001 §
Students	558 (30.9)	386 (31.5)	172 (29.6)	
Nurses	602 (33.3)	451 (36.8)	151 (25.9)	

Table 1. *Cont.*

Variable	Total (N = 1806)	TS (*n* = 1224; 94 Sessions)	NTS (*n* = 582; 39 Sessions)	*p*-Value
Physicians	222 (12.3)	200 (16.3)	22 (3.8)	
TCAI	150 (8.3)	74 (6.0)	76 (13.1)	
Other	274 (15.2)	113 (9.2)	161 (27.7)	
Institution type, *n* (%)				<0.001 §
Hospital	829 (54.8)	553 (55.6)	276 (53.2)	
University/community	685 (45.2)	442 (44.4)	243 (46.8)	

TS = technical skills; NTS = non-technical skills. † Age reported as median (n = 1547 participants (85.7%). ‡ Mann–Whitney U test. § Chi-square test. Percentages are calculated within each column.

3.2. Associations Between Participant Characteristics and Course Type

Spearman correlations are presented in Table 2. Age (ρ = 0.059, $p = 0.020$), gender (ρ = 0.065, $p = 0.020$), and professional category (ρ = 0.107, $p < 0.001$) were all significantly associated with course type. Institution type showed a significant chi-square but not a significant Spearman correlation (ρ = 0.026, $p = 0.308$), indicating the association is not ordinal in nature.

Table 2. Correlations between participant characteristics and course type (TS vs. NTS).

Participant Characteristic	ρ	*p*-Value	Interpretation
Age group	0.059	0.020	Younger: more TS
Gender	0.065	0.020	Female: more TS
Professional category	0.107	<0.001	"Other" category: more NTS
Institution type	0.026	0.308	Not significant (ordinal)

Positive ρ = stronger association with NTS courses; negative ρ = stronger association with TS courses. All correlations computed as Spearman ρ.

3.3. Perception Scores: Statistical and Practical Significance

Participants reported high satisfaction across all domains, with mean scores above 9.0 for both TS and NTS courses (Table 3). TS courses consistently received higher ratings than NTS courses across all four domains, and these differences reached statistical significance (all $p < 0.05$). However, Cohen's d effect size calculations revealed that all differences were small in magnitude (d range: 0.21 to 0.45).

Table 3. Perception scores by domain and course type, with effect sizes.

Domain	TS Mean (SD)	NTS Mean (SD)	*p*-Value	Cohen's d
Simulation/Methodology	9.59 (0.66)	9.23 (0.99)	<0.001	0.45
Teachers/Instructors	9.73 (0.56)	9.55 (0.79)	<0.001	0.28
Technological Materials	9.58 (0.75)	9.38 (0.90)	<0.001	0.22
Facilities/Technical Staff	9.66 (0.65)	9.51 (0.85)	<0.001	0.21
Overall perception	9.63 (0.60)	9.36 (0.84)	<0.001	0.38

TS = technical skills; NTS = non-technical skills; SD = standard deviation. Effect size benchmarks: small d = 0.20–0.49, medium d = 0.50–0.79, large d ≥ 0.80 (Cohen, 1988).

The largest effect was observed for the Simulation/Methodology subscale (TS: 9.59 ± 0.66 vs. NTS: 9.23 ± 0.99; d = 0.45, $p < 0.001$), suggesting that participants perceived the design and methodology of TS courses as somewhat more effective. The smallest effect was seen for the Facilities domain (TS: 9.66 ± 0.65 vs. NTS: 9.51 ± 0.85; d = 0.21, $p = 0.012$). The overall perception score was also significantly higher for TS (9.63 ± 0.60 vs. NTS: 9.36 ± 0.84; d = 0.38, $p < 0.001$). These differences are further illustrated in Figure 1.

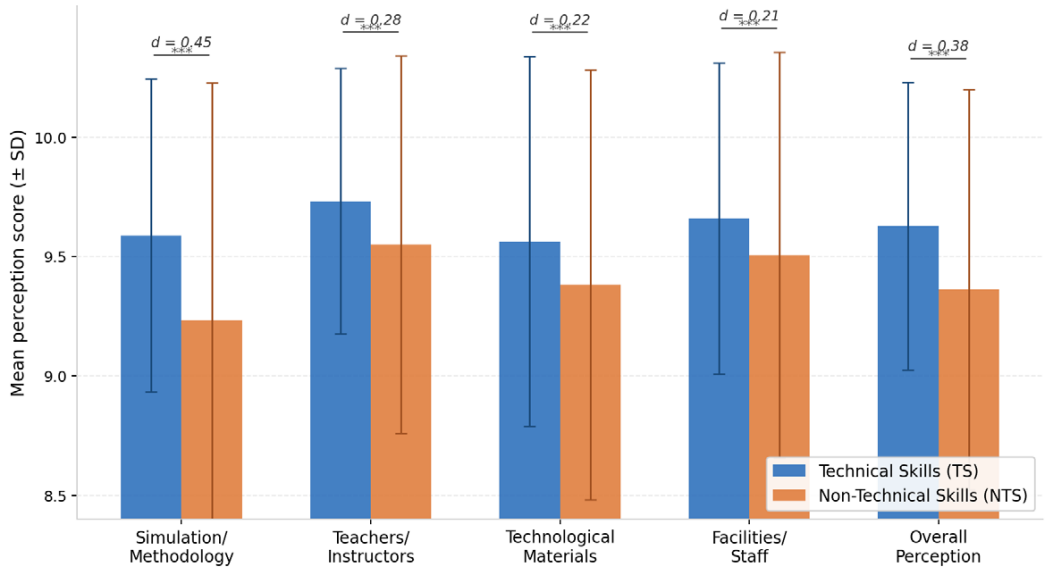

Figure 1. Mean perception scores (±SD) by domain and course type with Cohen's d effect sizes. *** $p < 0.001$ for all comparisons.

3.4. Item-Level Perception Analysis

The item-level analyses are presented in Table 4. The largest differences between TS and NTS courses were observed for P6 (Technical skills improved—d = 0.49), P8 (Error awareness—d = 0.43), and P5 (Knowledge improvement—d = 0.40), all belonging to the Simulation/Methodology subscale (all $p < 0.001$). Items related to instructors, facilities, and materials showed smaller, although still significant, differences (d = 0.17–0.29). Overall, the largest effect sizes were concentrated in items assessing perceived improvements in technical skills, knowledge acquisition, and error awareness, whereas items evaluating structural aspects of the training environment showed more limited variation between course types.

Table 4. Item-level perception scores by course type (TS vs. NTS).

Item	Description	TS M (SD)	NTS M (SD)	Cohen's d	*p*-Value
Simulation/Methodology					
P1	Session coordination	9.61 (0.72)	9.38 (0.99)	0.29	<0.001
P2	Activity timing	9.56 (0.82)	9.18 (1.23)	0.39	<0.001
P4	Met expectations	9.64 (0.73)	9.32 (1.09)	0.37	<0.001
P5	Knowledge improvement	9.54 (0.83)	9.12 (1.38)	0.40	<0.001
P6	Technical skills improved	9.57 (0.79)	9.02 (1.60)	0.49	<0.001
P7	Non-technical skills impr.	9.50 (0.88)	9.15 (1.31)	0.34	<0.001
P8	Error awareness	9.60 (0.75)	9.20 (1.23)	0.43	<0.001
P9	Realistic cases	9.66 (0.74)	9.49 (0.94)	0.22	<0.001
Teachers/Instructors					
P11	Safe learning environment	9.69 (0.67)	9.47 (0.96)	0.28	<0.001
P12	Instructor knowledge	9.77 (0.57)	9.62 (0.80)	0.23	<0.001
P13	Instructor clarity	9.75 (0.60)	9.56 (0.83)	0.27	<0.001
Tech Materials					
P3	Materials adequate	9.55 (0.87)	9.40 (0.93)	0.17	<0.001
P10	Realistic environment	9.57 (0.87)	9.37 (1.06)	0.22	<0.001

Table 4. *Cont.*

Item	Description	TS M (SD)	NTS M (SD)	Cohen's d	*p*-Value
Facilities/Staff					
P14	Facilities met expectations	9.62 (0.75)	9.45 (0.96)	0.21	<0.001
P15	Staff met needs	9.65 (0.71)	9.51 (0.90)	0.18	<0.001
P16	Would recommend CISTE	9.70 (0.65)	9.55 (0.92)	0.20	<0.001

TS = technical skills; NTS = non-technical skills; M = mean; SD = standard deviation. All *p*-values refer to independent-samples *t*-tests with Welch's correction. Items P3 and P10 belong to Tech Materials domain. Cohen's d: <0.20 = trivial; 0.20–0.49 = small; 0.50–0.79 = moderate. Note: Item P3 (Materials adequate) is listed within the Tech Materials domain section alongside P10, in accordance with its domain assignment. Items are grouped by domain rather than by item number; all 16 items are present and accounted for.

3.5. Perception by Professional Category

Table 5 presents overall perception scores stratified by professional category and course type. Physicians showed the highest overall satisfaction in TS courses (9.79 ± 0.39) and the largest effect size between course types (d = 0.78, $p = 0.029$). Nurses showed the largest n with a moderate effect (d = 0.61, $p < 0.001$). The "Other" category was the only group showing no significant difference between TS and NTS (d = 0.20, $p = 0.093$), consistent with their predominant participation in NTS courses. A Kruskal–Wallis test across professional categories confirmed significant overall differences in perception ($p < 0.001$). Overall perception scores by professional category are displayed in Figure 2.

Table 5. Overall perception scores by professional category and course type.

Professional Category	n TS	TS M (SD)	n NTS	NTS M (SD)	Cohen's d
Nurse	451	9.63 (0.59)	151	9.20 (1.00)	0.61 ($p < 0.001$)
Physician	200	9.79 (0.39)	22	9.46 (0.65)	0.78 ($p = 0.029$)
TCAI	74	9.58 (0.56)	76	9.30 (0.80)	0.41 ($p = 0.014$)
Other	113	9.68 (0.53)	161	9.56 (0.61)	0.20 ($p = 0.093$)
Student	386	9.53 (0.71)	172	9.34 (0.87)	0.26 ($p = 0.010$)

TS = technical skills; NTS = non-technical skills; M = mean; SD = standard deviation. Cohen's d and *p*-values from Welch's *t*-test. "Other" $p = 0.093$ (not significant). Overall perception = mean of all 16 items.

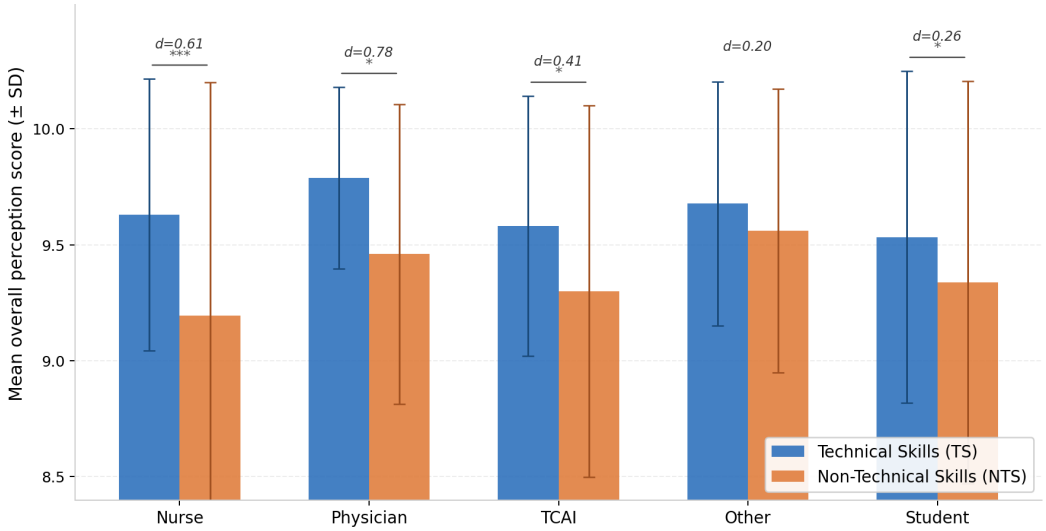

Figure 2. Overall perception scores (±SD) by professional category and course type. * $p < 0.05$; *** $p < 0.001$.

4. Discussion

This study analyzed healthcare professionals' perceptions of simulation-based training classified by primary learning objective, technical skills (TS) versus non-technical skills

(NTS), using Kirkpatrick's Level 1 evaluation framework, and examined associations with participant characteristics at an SSH-accredited simulation center. Overall, participants reported excellent perceptions across all domains (all means > 9.0 on a 10-point scale), consistent with prior studies reporting high learner satisfaction in high-fidelity simulation contexts (Bdiri Gabbouj et al., 2024; McGaghie et al., 2010).

A key finding of this study is that, although differences were identified between TS and NTS courses across all perception domains, these differences were generally limited in magnitude. This suggests that, despite some variation in how participants evaluated the two types of courses, perceptions were broadly comparable across training contexts, highlighting more similarities than substantial disparities in the overall learning experience. Consequently, the presence of statistical significance alone does not imply that these differences are educationally meaningful. The EFA establishes that the questionnaire measures a single general satisfaction construct, so domain-level effect sizes represent facets of this construct rather than independent dimensions. Taken together, the data indicate that high-fidelity simulation training is perceived as highly effective regardless of course type. The item-level pattern reveals that the TS advantage concentrates specifically in items measuring concrete and observable learning gains (P5, P6, P7, P8), while structural items, instructors, facilities, and materials show minimal differences, a theoretically coherent and practically actionable finding.

The slightly higher satisfaction scores for TS courses likely reflect characteristics intrinsic to procedural training: TS sessions offer hands-on skill practice with immediate and tangible performance feedback, clear progression cues, and measurable outcomes (e.g., successful completion of a procedure). In contrast, NTS training, while addressing critical competencies such as communication, leadership, and crisis resource management, involves more abstract learning objectives and less immediate feedback, which may contribute to somewhat lower satisfaction scores even when educational quality is high (Cloonan et al., 2026; Cheng et al., 2014). This interpretation is consistent with findings from Peddle (Peddle, 2019) and others who noted that participants in NTS-focused simulation may perceive their learning as less concrete.

Gender was not significantly associated with course type or perception scores in this study, consistent with evidence from trauma team simulation research showing minimal impact of gender on NTS learning (Ylönen et al., 2025). This null finding is nonetheless informative, as it suggests that gender equity in access to simulation training, at least within this center's programming, is being maintained across both course types.

An interesting finding was that the "Other" professional category did not show clear differences in the evaluation of TS and NTS courses. This group includes professionals such as administrative staff, educators, security personnel, psychologists, and social workers, whose daily work is primarily centred on communication, coordination, and interpersonal interaction rather than procedural clinical tasks. In this context, non-technical competencies may be perceived as directly connected to their routine professional practice, which could contribute to a more similar evaluation of both training formats. In contrast, clinically trained professionals appeared to value TS courses more positively, suggesting that perceptions of simulation training may be influenced by the extent to which the competencies addressed align with the core activities of each professional role. From a theoretical perspective, the Kirkpatrick Level 1 data presented here are foundational but insufficient for a full evaluation of simulation program effectiveness. High learner satisfaction is a necessary but not sufficient condition for learning transfer (Kirkpatrick & Kirkpatrick, 2016; Bdiri Gabbouj et al., 2024). Future research at this center should examine Level 2 outcomes (knowledge and skill acquisition), Level 3 outcomes (behavior change in clinical practice),

and ideally Level 4 outcomes (patient safety and clinical performance indicators). The consistently high Level 1 scores provide a favorable starting point for these investigations.

The findings carry several practical implications for simulation program design, policy, and educational practice.

A further reflection emerging from these results concerns the conceptual framing of non-technical skills themselves. The prevailing label, "non-technical" or "soft" skills, implies a secondary or supplementary nature relative to procedural competencies. This framing is, however, epistemologically misleading: NTS encompasses the most structurally complex and cognitively demanding competencies required in clinical practice, including situation awareness, leadership, crisis resource management, and interprofessional communication. Their difficulty lies precisely in the fact that their acquisition is not linear, their progress is less observable in the short term, and their measurement requires validated instruments that the literature itself acknowledges as still under development (Gawronski et al., 2022). Abildgren et al. (2022) demonstrated in a systematic review published in Advances in Simulation that human factor skills, the broader category that encompasses NTS, are trainable to a degree comparable to technical skills, but crucially, evidence on their retention and transfer to clinical practice remains insufficient (Abildgren et al., 2022). This gap between the recognized importance of NTS and the methodological tools available to capture their impact is a structural problem in health professions education, not a reflection of their intrinsic value.

This structural complexity has direct implications for interpreting learner satisfaction data. Research consistently links NTS failures to a disproportionate share of adverse clinical events: Alexandrino et al. (2023) noted in a narrative review in Frontiers in Medicine that root cause analyses of surgical adverse events identify failures in coordination, planning, and communication, all NTS domains, as the dominant causes of medical error, ahead of purely technical failures (Alexandrino et al., 2023). Likewise, Tam et al. (2024), in a systematic review in Surgical Endoscopy, found that acute stress degrades NTS even more markedly than technical skills, underscoring their central role in high-stakes clinical performance (Tam et al., 2024). Taken together, this evidence supports the interpretation that the perceptual gap observed in the present study does not signal lower educational quality of NTS courses, but rather reflects the inherent difficulty of making abstract, behavioural learning gains immediately visible to learners. This interpretation is consistent with prior work by Redjem et al. (2025), who showed in a systematic review in Nurse Education Today that NTS training evaluations rarely exceed Kirkpatrick Level 2, not because NTS training is less effective, but because its higher-level impacts are harder to measure and communicate (Redjem et al., 2025). Strengthening NTS debriefing with explicit behavioural metrics and anchored performance feedback, rather than general reflective discussion, is therefore a priority for simulation educators seeking to close this perceptual gap, as also supported by recent Spanish-context evidence on nursing simulation (García-Salido et al., 2024). The item-level data provide empirical support for this recommendation. Item P8 ("Error awareness") showed one of the largest differences in favour of TS courses ($d = 0.43$), a pattern that is consistent with the nature of procedural training, where errors tend to be discrete, observable, and immediately identifiable (e.g., an incorrect catheter placement or a medication dosing error). By contrast, in NTS scenarios, errors are often more diffuse, emerging at the level of communication patterns or team interaction, and may be interpreted differently depending on the participant's role within the scenario. In the absence of structured debriefing approaches that make these aspects explicit, such as anchored behavioural rating scales or guided team reflection, these types of errors are less likely to be consistently recognised. This finding therefore points to error awareness as a

key area for strengthening the instructional design of NTS training, particularly through more structured and standardised debriefing strategies.

This study makes several contributions to the existing literature on simulation-based health professions education. First, to our knowledge, it is one of the few studies to systematically compare Kirkpatrick Level 1 outcomes between TS and NTS courses in an SSH-accredited multi-institutional centre, using a large, real-world sample (n = 1806) and effect size analysis to move beyond statistical significance. Most prior studies evaluating NTS simulation have focused on specific scenarios or specialties (e.g., operating room, emergency department), typically with smaller samples and without direct TS versus NTS comparison at the programme level. Second, the inclusion of participants across multiple professional categories, nurses, physicians, TCAI, students, and other healthcare roles, provides a more representative picture of real-world simulation populations than studies restricted to a single profession. Third, the use of Cohen's d across all domains and items, combined with the explicit distinction between statistical and educational significance, is methodologically rigorous and directly applicable to curriculum planning decisions. Fourth, the study was conducted in an accredited centre operating under SSH and INACSL standards, enhancing the generalizability of findings to similarly resourced simulation environments. Together, these contributions position the present study as a meaningful empirical reference for simulation programme evaluation in the European health professions education context.

4.1. Implications for Practice

First, the small effect sizes suggest that NTS training should not be perceived as educationally inferior to TS training; its slightly lower satisfaction scores likely reflect inherent characteristics of the learning content rather than deficiencies in program quality. Simulation center managers and curriculum designers should communicate this distinction clearly to stakeholders and prioritize the continued development of NTS offerings. Investment in deliberate instructional strategies specific to NTS, such as structured debriefing with explicit performance metrics, scenario complexity calibration matched to career stage, and structured peer feedback mechanisms, may help close the perceptual gap and maximize the educational impact of these courses (Cloonan et al., 2026; Cheng et al., 2014).

Second, the career-stage distribution of participants, with younger students and trainees engaging primarily in TS courses and more experienced professionals gravitating toward NTS, has direct implications for program scheduling and targeted recruitment. Simulation centers should consider developing career-stage-specific simulation pathways that systematically introduce NTS training earlier in professional development, rather than treating it as a later-career activity. In this respect, the student subgroup warrants particular attention. Although students mainly participated in TS courses, they exhibited a relatively small difference in satisfaction between TS and NTS formats, especially when compared with physicians and nurses, who showed a much stronger preference for TS. While this pattern has been interpreted in relation to students' early stage of professional development, it may also point to a broader shift in educational expectations. Current health sciences curricula place increasing emphasis on interprofessional learning, teamwork, and communication from the outset of training, which may foster a more balanced appreciation of both technical and non-technical competencies. If so, students may enter simulation-based education with greater openness to NTS approaches than more experienced clinicians. This possibility highlights an important area for future research on how educational exposure shapes perceptions of different simulation modalities over time.

Third, the consistently high instructor ratings across both course types (mean scores ≥ 9.51) highlight instructor expertise as a key driver of learner satisfaction regardless of

course content (Bdiri Gabbouj et al., 2024; McGaghie et al., 2010). This finding supports sustained investment in instructor training and development programs, particularly in preparation for facilitating the more complex interpersonal dynamics that characterize NTS-focused simulation. Centers that cannot yet demonstrate high instructor ratings for NTS courses may consider a phased implementation model, beginning with intensive instructor preparation before expanding the NTS portfolio.

4.2. Limitations

This study has several limitations. First, data were collected through a self-reported questionnaire, which may be subject to response and social desirability biases. Second, the cross-sectional retrospective design limits causal inference. Third, the study was conducted at a single accredited simulation center, which may limit generalizability to settings with different resources, institutional cultures, or accreditation standards. Fourth, the exploratory factor analysis conducted on the complete-case item data (N = 1814) supported a unidimensional structure of the instrument (KMO = 0.957; eigenvalue = 10.143; 63.4% of variance explained; item loadings 0.49–0.85). Consistent with this result, the overall perception score should be considered the primary psychometrically supported outcome, while domain scores are better understood as exploratory groupings of items within a single underlying construct of satisfaction. In line with this interpretation, the Fornell–Larcker criterion was not satisfied for any domain, indicating substantial overlap between subscales and reinforcing the absence of empirically distinct dimensions. Although this limits the extent to which domain-level differences can be interpreted as independent constructs, future work should include confirmatory factor analyses in independent samples and comparisons with externally validated instruments to further strengthen the measurement model. Fifth, this study examined only Kirkpatrick Level 1 (learner reaction) outcomes; learning acquisition, behavior transfer, and patient safety impacts were not assessed.

5. Conclusions

Simulation-based training at this SSH-accredited, multi-institutional clinical center is highly valued by healthcare professionals for both technical and non-technical skills. While TS courses were consistently rated higher across all perception domains, effect sizes ranged from small to moderate-approaching (d = 0.21–0.45), indicating that the practical educational significance of these differences is limited. Both course types achieved excellent learner satisfaction, supporting the quality and effectiveness of the simulation program as a whole.

Participant characteristics, particularly age, professional category, and institution type, are significantly associated with the type of simulation training attended, consistent with a career-stage model of professional development. These associations have actionable implications for curriculum design, learner recruitment, and institutional planning.

Future work should prioritize Kirkpatrick Level 2–4 outcome evaluations, longitudinal follow-up of skill transfer to clinical practice, and formal psychometric validation of the perception instrument used here. The consistently positive Level 1 findings reported in this study provide a strong empirical foundation for these next steps.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci16050807/s1, Table S1: Exploratory Factor Analysis: Factor Loadings, Reliability, and Validity Indices for the 16-Item Perception Questionnaire (Leighton et al., 2015; Tabachnick & Fidell, 2019).

Author Contributions: S.R.-V., L.G.-C., J.M.E.B., J.B.R., R.P., M.A.-L. and G.M.-L. contributed to this work as follows: Conceptualization: S.R.-V., R.P., M.A.-L. and G.M.-L.; Methodology: S.R.-V., L.G.-C.,

J.M.E.B. and G.M.-L.; Formal analysis: S.R.-V., L.G.-C., R.P. and G.M.-L.; Investigation/Data collection: J.M.E.B., J.B.R. and L.G.-C.; Project administration and resources: S.R.-V., M.A.-L., L.G.-C. and G.M.-L.; Data curation and software/visualization: S.R.-V., L.G.-C. and G.M.-L.; Writing—Original Draft: S.R.-V., L.G.-C. and G.M.-L.; Writing—Review & Editing: All authors; Supervision: S.R.-V., R.P. and G.M.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially supported by a 2026 grant programme within the framework of the 2024–2027 collaboration agreement between the Diputació de Tarragona and the Universitat Rovira i Virgili (URV), and supported by Grup de Recerca en Infermeria Avançada (CARING) under the SGR-Cat 2021 call (2021 SGR 00161).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Research Ethics Committee on Medicinal Products (Comitè d'Ètica d'Investigació amb Productes Sanitàris i de la Salut d'Atenció Primària, CEIPSA) on 12 Jan 2023 under project code 2023 TDO 0112.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The anonymised dataset supporting the findings of this study and the questionnaire instrument are available as Supplementary Materials accompanying this article. Raw data files are deposited at the institutional repository of Universitat Rovira i Virgili (https://repositori.urv.cat (accessed on 12 May 2026)) and will be made publicly accessible upon acceptance. Statistical analysis scripts (SPSS syntax) are available from the corresponding author upon reasonable request.

Acknowledgments: We thank all healthcare professionals and institutions who participated in simulation-based training at the Centre d'Innovació i Simulació Territorial (CISTE) during the study period. During the preparation of this manuscript, Claude (Anthropic) was used to assist with structural organisation and language editing of manuscript sections, conversion of citation formats, and generation of data visualisations (Figures 1 and 2). All analytical decisions, interpretation of results, and scientific conclusions were made exclusively by the authors. AI-generated text was reviewed, edited, and validated by the authors prior to submission.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ALS	Advanced Life Support
AVE	Average Variance Extracted
CEIPSA	Comitè d'Ètica d'Investigació en Persones, Societat i Medi Ambient
CFA	Confirmatory Factor Analysis
CISTE	Centre d'Innovació i Simulació Territorial
CRM	Crisis Resource Management
EFA	Exploratory Factor Analysis
INACSL	International Nursing Association for Clinical Simulation and Learning
KMO	Kaiser–Meyer–Olkin measure of sampling adequacy
ML	Maximum Likelihood
NTS	Non-Technical Skills
PA	Parallel Analysis
SBE	Simulation-Based Education
SD	Standard Deviation
SPSS	Statistical Package for the Social Sciences
SSH	Society for Simulation in Healthcare
TCAI	Tècnic en Cures Auxiliars d'Infermeria
TS	Technical Skills
URV	Universitat Rovira i Virgili

References

Abildgren, L., Lebahn-Hadidi, M., Mogensen, C. B., Toft, P., Nielsen, A. B., Frandsen, T. F., Steffensen, S. V., & Hounsgaard, L. (2022). The effectiveness of improving healthcare teams' human factor skills using simulation-based training: A systematic review. *Advances in Simulation*, *7*, 12. [CrossRef]

Alexandrino, H., Martinho, B., Ferreira, L., & Baptista, S. (2023). Non-technical skills and teamwork in trauma: From the emergency department to the operating room. *Frontiers in Medicine*, *10*, 1319990. [CrossRef]

Bdiri Gabbouj, S., Zedini, C., & Naija, W. (2024). Nursing students' satisfaction and self-confidence with simulation-based learning and its associations with simulation design characteristics and educational practices. *Advances in Medical Education and Practice*, *15*, 1093–1102. [CrossRef]

Cheng, A., Eppich, W., Grant, V., Sherbino, J., Zendejas, B., & Cook, D. A. (2014). Debriefing for technology-enhanced simulation: A systematic review and meta-analysis. *Medical Education*, *48*(7), 657–666. [CrossRef]

Cloonan, M. R., Maginot, E. R., Tieken, K. R., Merani, S., Schindler, N., Tanner, T. N., & Fingeret, A. L. (2026). Evaluating patterns in technical and nontechnical skills feedback for surgical residents: Insights from SIMPL app data. *Global Surgical Education*, *5*, 3. [CrossRef]

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. [CrossRef] [PubMed]

Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N., & Patey, R. (2003). Anaesthetists' non-technical skills (ANTS): Evaluation of a behavioural marker system. *British Journal of Anaesthesia*, *90*(5), 580–588. [CrossRef] [PubMed]

Flin, R., Martin, L., Goeters, K., Hoermann, J., Amalberti, R., Valot, C., & Nijhuis, H. (2003). Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. *Human Factors and Aerospace Safety*, *3*, 95–117.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. [CrossRef]

García-Salido, A., Montilla-Aparicio, M. J., García-Teresa, M. Á., Angulo-Fraile, E., & Giménez-Palop, O. (2024). Non-technical skills training in nurses through in-situ simulation: A pilot study. *Enfermería Intensiva*, *35*(3), 215–224.

Gawronski, O., Thekkan, K. R., Genna, C., Egman, S., Sansone, V., Erba, I., Vittori, A., Varano, C., Dall'Oglio, I., Tiozzo, E., & Chiusolo, F. (2022). Instruments to evaluate non-technical skills during high fidelity simulation: A systematic review. *Frontiers in Medicine*, *9*, 986296. [CrossRef]

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. [CrossRef]

Kim, J., Neilipovitz, D., Cardinal, P., & Chiu, M. (2009). A comparison of global rating scale and checklist scores in the validation of an evaluation tool to assess performance in the resuscitation of critically ill patients during simulated emergencies. *Simulation in Healthcare*, *4*(1), 6–16. [CrossRef]

Kim, J., Neilipovitz, D., Cardinal, P., Chiu, M., & Clinch, J. (2006). A pilot study using high-fidelity simulation to formally evaluate performance in the resuscitation of critically ill patients: The university of Ottawa critical care medicine, high-fidelity simulation, and crisis resource management I study. *Critical Care Medicine*, *34*(8), 2167–2174. [CrossRef] [PubMed]

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation*. Association for Talent Development.

Leighton, K., Ravert, P., Mudra, V., & Macintosh, C. (2015). Updating the Simulation Design Scale: An instrument to evaluate simulation design features. *Journal of Nursing Education*, *54*, S58–S65. [CrossRef]

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Education*, *44*(1), 50–63. [CrossRef]

Muñoz Gualán, G. G., & Sierra, R. E. (2025). La simulación clínica en la educación médica moderna: Revisión de revisiones. *Educación Médica*, *19*(1), 102–116. [CrossRef]

Patten, E., Miller, C., Wong, P. T., Cochrane, L., & Jarden, R. (2026). Interprofessional perioperative simulation team training non-technical skill measurement and outcomes: A scoping review. *Journal of Interprofessional Care*, *40*, 344–364. [CrossRef] [PubMed]

Peddle, M. (2019). Participant perceptions of virtual simulation to develop non-technical skills in health professionals. *Journal of Research in Nursing*, *24*(3–4), 167–180. [CrossRef]

Pimenta, I. D. S. F., Mata, Á. N. d. S., Bezerra, I. N. M., de Araújo, R. M. P., Santana, J. L. A., Capucho, H. C., de Souza-Oliveira, A. C., Echevarría-Pérez, P., Guillén-Martínez, D., & Piuvezam, G. (2025). Effectiveness of non-technical skills training in intensive care units: A systematic review. *BMC Medical Education*, *25*, 847. [CrossRef]

Prete, D., Tamburri, L., Rolston, N., Sturgill, M., & Bridgeman, M. (2024). Using simulation to introduce students and healthcare professionals to losses experienced by older adults. *BMC Medical Education*, *24*, 96. [CrossRef]

Pucer, P., Martinović, K., Karnjuš, I., & Renko, J. (2025). Development and improvement of non-technical skills in interprofessional healthcare teams through simulation-based experiences: A systematic review. *Clinical Simulation in Nursing*, *102*, 101726.

Redjem, I. D., Huaulmé, A., Jannin, P., & Michinov, E. (2025). Crisis management in the operating room: A systematic review of simulation training to develop non-technical skills. *Nurse Education Today*, *147*, 106583. [CrossRef] [PubMed]

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*, 1273–1296. [CrossRef]

Tam, A., Bateman, S., Buckingham, G., Wilson, M., Melendez-Torres, G. J., Vine, S., & Clark, J. (2024). The effects of stress on surgical performance: A systematic review. *Surgical Endoscopy*, *39*(1), 77–98. [CrossRef] [PubMed]

World Medical Association. (2013). Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191–2194. [CrossRef]

Ylönen, M., Heinänen, M., Lauritsalo, S., Paloneva, J., & Rosqvist, E. (2025). Influence of trainee's demographic variables on non-technical skill learning after trauma team simulation: Cross-sectional study. *BMC Medical Education*, *25*, 1107. [CrossRef]

Zhang, C. (2023). A literature study of medical simulations for non-technical skills training in emergency medicine: Twenty years of progress, an integrated research framework, and future research avenues. *International Journal of Environmental Research and Public Health*, *20*(5), 4487. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.