The Canadian Journal of Neurological Sciences (2026), **53**, 739–748
doi:10.1017/cjn.2025.10404

Canadian Journal of Neurological Sciences
Journal Canadien des Sciences Neurologiques

Original Article

Face, Content, Construct and Convergent Validity of a Surgical Spine Simulator for Pedicle Screw Insertions

Trisha Tee[1,2], Noel Abboud[1,2], Bilal Tarabay[1,2], Abdulmajeed Albeloushi[1,3,4], Puja Pachchigar[1,2], Mohamed Alhantoobi[1,2,5,6], Nour Abou Hamdan[1,2], Recai Yilmaz[1,2], Ali Fazlollahi[1,2] and Rolando F. Del Maestro[1,2,3]

[1]Neurosurgical Simulation and Artificial Intelligence Learning Centre, Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada, [2]Faculty of Medicine and Health Sciences, Department of Experimental Surgery, McGill University, Montreal, Quebec, Canada, [3]Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada, [4]Department of Neurosurgery, Ibn Sina Hospital, Ministry of Health, Kuwait City, Kuwait, [5]Department of Neurosurgery, Hamilton General Hospital, McMaster University Medical Centre, Hamilton, Ontario, Canada and [6]Department of Neurosurgery, Zayed Military Hospital, Abu Dhabi, United Arab Emirates

ABSTRACT: ***Background:*** Spine simulators offer learners risk-free environments to develop psychomotor skills for pedicle screw insertions. The virtual reality TSYM simulator deconstructs and simulates pedicle screw insertions. This case series study investigates face, content, construct, and convergent validity of an L4–L5 bilateral pedicle screw insertion on the TSYM simulator. ***Methods:*** Neurosurgical-orthopedic residents, fellows, and spine surgeons performed an L4–L5 bilateral pedicle screw insertion on the TSYM simulator. Participants were classified a priori into skilled (postgraduate year (PGY) 5–6, fellows, and consultant neurosurgeons or orthopedic surgeons) or less skilled (PGY 1–4) groups. Face and content validity were assessed utilizing a 7-point Likert scale. Construct validity was determined by investigating group differences in simulation-derived performance metrics and the Objective Structured Assessment of Technical Skills (OSATS) ratings. Convergent validity was examined by correlating simulation-derived performance metrics and OSATS ratings. ***Results:*** Thirteen skilled and 14 less skilled participants were included in this study. Eight of nine face and content validity statements were rated a median ≥ 4. Significant differences between the groups were found for four simulation-derived performance metrics ($P < 0.05$) and all OSATS categories ($P < 0.001$). Three simulation-derived performance metrics (maximum force and tool contact using the simulated screwdriver and three-dimensional velocity using the tap) significantly correlated with OSATS ratings. ***Conclusion:*** The L4–L5 bilateral pedicle screw insertion simulation on the TSYM platform demonstrated mixed and variable evidence for face, content, construct and convergent validity, supporting its educational potential for spine surgery training, but improvements are needed to optimize learning.

RÉSUMÉ : Validité apparente, validité de contenu, validité conceptuelle et validité convergente d'un simulateur de chirurgie de la colonne vertébrale pour l'insertion de vis pédiculaires. ***Contexte :*** Les simulateurs de chirurgie de la colonne vertébrale offrent aux apprenants un environnement sans risque pour accroître leurs compétences psychomotrices en matière d'insertion de vis pédiculaires. Le simulateur de réalité virtuelle TSYM permet de décomposer et de simuler l'insertion de vis pédiculaires. Cette étude de série de cas entend examiner la validité apparente, la validité de contenu, la validité conceptuelle et la validité convergente d'une insertion bilatérale de vis pédiculaires en L4 et L5 au moyen du simulateur TSYM. ***Méthodes :*** Des résidents en neurochirurgie et en orthopédie, des boursiers et des chirurgiens de la colonne vertébrale ont réalisé une insertion bilatérale de vis pédiculaires en L4 et L5 à l'aide du simulateur TSYM. Les participants ont été classés a priori en deux groupes : ceux étant expérimentés (des étudiants en 5e et 6e années d'études supérieures, des boursiers et des neurochirurgiens ou bien des chirurgiens orthopédistes consultants) et ceux étant moins expérimentés (des étudiants ayant entre une et quatre années d'études supérieures). La validité apparente et la validité de contenu ont été évaluées à l'aide d'une échelle de Likert à 7 points. La validité conceptuelle, quant à elle, a été déterminée en examinant les différences entre les groupes, et ce, en se basant sur les mesures de performance lors d'une simulation et sur les notes de l'évaluation objective structurée des compétences techniques (« OSATS » en anglais). Enfin, la validité convergente a été examinée en corrélant les mesures de performance lors d'une simulation et les notes à l'OSATS. ***Résultats :*** Au total, 13 participants qualifiés et 14 participants moins qualifiés ont été inclus dans cette étude. Huit des neuf énoncés de validité apparente et de validité de contenu ont obtenu une note médiane de ≥ 4. Des différences significatives entre les groupes ont été observées pour quatre mesures de performance lors d'une simulation ($p < 0{,}05$) et toutes les catégories de l'OSATS ($p < 0{,}001$). Trois mesures de performance lors d'une simulation (force maximale et contact avec l'outil à l'aide du tournevis simulé ; vitesse 3D à l'aide du taraud) étaient significativement corrélées aux notes à l'OSATS. ***Conclusion :*** La simulation d'insertion bilatérale de vis pédiculaires en L4 et L5 au moyen du simulateur TSYM a démontré des preuves mitigées et variables en matière de validité apparente, de validité de contenu, de validité conceptuelle et de validité convergente, ce qui, en dépit d'améliorations nécessaires pour optimiser l'apprentissage, confirme son potentiel éducatif pour la formation en chirurgie de la colonne vertébrale.

Keywords: Construct validity; pedicle screw insertion; virtual reality simulation; surgical simulation training; surgical education

(Received 7 April 2025; final revisions submitted 27 July 2025; date of acceptance 10 August 2025; First Published online 26 August 2025)

Corresponding author: Trisha Tee; Email: trisha.tee@mail.mcgill.ca

Cite this article: Tee T, Abboud N, Tarabay B, Albeloushi A, Pachchigar P, Alhantoobi M, Abou Hamdan N, Yilmaz R, Fazlollahi A, and Del Maestro RF. (2026) Face, Content, Construct and Convergent Validity of a Surgical Spine Simulator for Pedicle Screw Insertions. *The Canadian Journal of Neurological Sciences* **53**: 739–748, https://doi.org/10.1017/cjn.2025.10404

© The Author(s), 2025. Published by Cambridge University Press on behalf of Canadian Neurological Sciences Federation. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:23:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cjn.2025.10404

Highlights

- Provides evidence of validity for a virtual reality spine simulator's L4–L5 pedicle screw insertion scenario.
- Utilizes a comprehensive validation approach using traditional (face, content, construct and convergent validity) and contemporary validity frameworks.
- Suggests that combining simulator-derived metrics with OSATS ratings can enhance our understanding and assessment of surgical skills.

Introduction

Surgical training involves balancing skill acquisition with ensuring patient safety.[1–3] This becomes particularly relevant in spine surgery due to its complexity and variability in resident exposure.[4–6] Pedicle screw insertion is a common but technically demanding spine surgical procedure, involving a steep learning curve.[6,7] Potential risks include malposition rates ranging between 4.2% and 7.8%, making acquiring proficiency under direct supervision essential.[8–11]

Virtual reality simulation offers a promising role in providing a risk-free environment for procedural learning and skill refinement.[7,12,13] However, current spine surgery simulators often lack high fidelity and comprehensive validity.[14–17] In a recent review of augmented reality, virtual reality and mixed reality related to learning in healthcare professions, only a small fraction of training tools had evidence supporting face, content or construct validity.[18] The study highlights the need for face, content and construct validity assessment and the development of more relevant simulation training tools.[14–18]

The TSYM Symgery virtual reality platform aims to fill this gap by providing realistic pedicle insertion and performance feedback.[19–21] This study evaluates its educational utility through established traditional and contemporary validation frameworks.[22–24] Specifically, it assesses face and content validity via questionnaire responses from experts, and construct validity by comparing simulator performance metrics and Objective Structured Assessment of Technical Skills (OSATS) scores between "less skilled" and "skilled" groups.[22,25,26] Convergent validity is explored by correlating simulator performance metrics with OSATS scores, the gold standard in surgical assessment.[24,27–29]

This study seeks to answer the research question: What evidence of validity supports the educational utility of the TSYM simulator for spine surgery training? Therefore, the objectives of this case series study were 1) to evaluate face and content validity for an L4–L5 bilateral pedicle screw insertion simulation on the TSYM simulator platform, 2) to use simulation-derived metrics and the assessment of simulated pedicle screw insertion operative performance utilizing OSATS to assess construct validity, 3) to establish convergent validity employing simulation-derived metrics and simulated pedicle screw insertion operative performance OSATS and 4) to attempt to use the results to construct an argument supporting the TSYM simulator's use for training residents and fellows in the L4–L5 bilateral pedicle screw insertion.

Methods

Participants

Neurosurgical and orthopedic residents, spine fellows, nonspine neurosurgical fellows who had experience in pedicle screw insertion and neurosurgical and orthopedic spine surgeons participated in this case series study. An exclusion criterion was previous experience with the TSYM simulator. Based on information from orthopedic and neurosurgical training programs in Quebec universities related to resident experience with clinical pedicle screw insertions, participants were categorized a priori into two groups: skilled participants (postgraduate year (PGY) 5–6 residents, fellows and spine surgeons) and less skilled residents in PGY 1 to 4. Participants signed an informed consent approved by the Neurosciences-Psychiatry McGill University Health Center Research Ethics Board. Participants then completed a demographic questionnaire and were provided with standardized written and verbal instructions on the steps and instruments to complete the simulated L4–L5 bilateral pedicle screw insertion on the TSYM simulator. Participants first performed a dry lab and an L2 simulated laminectomy procedure to become acquainted with the TSYM simulator (see Supplemental Information). After completing these tasks, participants performed a simulated L4–L5 bilateral pedicle screw insertion on the TSYM simulator. No time limit was imposed, but each step was dependent, requiring participant confirmation of step completion before proceeding. This article follows the Strengthening the Reporting of Observational Studies in Epidemiology reporting guidelines.[30]

Virtual reality simulator platform

The TSYM Symgery simulation platform, developed by Cedarome Canada Inc. dba Symgery. (Montreal, Canada), was utilized in this study (Figure 1A). This simulator's three-dimensional (3D) intraoperative spinal surgical procedures rely on a voxel-based system [26] (Figure 1B). The simulator consists of a single haptic arm that provides continuous tactile, auditory and visual feedback while using the simulator's surgical instruments (Figure 1C). This system is equipped with pre-programed surgical tools and captures multiple performance metrics, enabling a detailed analysis of surgical performance. The pedicle screw insertion simulation task consists of one animated and four deconstructed interactive steps described in Table 1. These steps were repeated for each screw. For standardization purposes, users performed the pedicle screw insertions using constant magnification and inserted 6.5 × 45 mm pedicle screws in a predetermined order: left L5, left L4, right L5, right L4 (see Supplementary Information). Participants had access to live X-rays to verify the entry point and angles for pedicle cannulation and confirm the accuracy of inserted screws. The Supplementary Video shows a skilled participant performing a pedicle screw insertion on the simulator.

Face and content validity

The spine surgeons and fellows assessed the face and content validity of the pedicle screw insertion simulation using questionnaires assessed with a 7-point Likert scale with 1 being "completely unrealistic" and 7 being "completely realistic".[26,31] While there is no universal median value for establishing sufficient face and content validity, this study considered the overall simulated procedure and its deconstructed tasks to demonstrate such validity if questionnaires achieved a median $\geq$ 4.0 on the 7-point Likert scale, consistent with prior studies.[26,31]

Construct validity

To assess construct validity, the study assessed each pedicle screw insertion independently and employed performance metrics derived from the TSYM simulator and blinded expert scoring using OSATS.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:23:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cjn.2025.10404



Figure 1. TSYM virtual reality simulator platform developed by Cedarome Canada Inc. dba Symgery (Montreal, Canada) (A) The TSYM simulator set up, showing the (1) robotic arm that uses and provides advanced haptic feedback technology, (2) the different tool handles that can be used in the simulated scenario, (3) 3D monitor, (4) pedals for activating fluoroscopy and (5) secondary monitor. (B) A neurosurgical resident performing a task on the simulator, demonstrating its practical use in a training scenario. (C) The tool handles are available to mimic an array of tools in the virtual environment.

Simulation-derived tool metrics

The TSYM simulator continuously assessed several features of performance during pedicle screw insertion. Data on each tool's 3D velocity, 3D force, maximum force, 3D acceleration and tool tissue contact were collected for each screw. The 3D force and maximum force refer to the forces applied to the haptic arm while using the tool. The 3D velocity and 3D acceleration of each tool are derived from the position of the tool's tip in space. The tools that were assessed can be found in Table 1. The rationale to treat each pedicle screw insertion by each participant independently was that each screw insertion involved a different simulated vertebral entry point, orientation, and angulation.

Randomized-blinded OSATS assessment

In concert with the simulator-derived performance metrics, the study utilized the validated methodology of learner-operative performance assessment employed by surgical educators in human operative settings, OSATS ratings, to determine construct validity.[28,29] Each participant's simulated L4–L5 bilateral pedicle screw insertion was recorded on-screen, which was later subdivided into four videos, one for each pedicle screw insertion. Video recordings of each lumbar pedicle screw insertion were randomized and blindly rated by two experts with experience performing human pedicle screw insertions. The OSATS scale was adapted to the simulator's capabilities, resulting in five items (respect for tissue, instrument handling, the economy of movement, flow and knowledge of procedure) and an overall rating. Each performance was rated on a 7-point Likert scale. The OSATS scale demonstrated excellent internal consistency ($\alpha = 0.97$ [95% CI, 0.96, 0.98]) and excellent inter-rater reliability ($\alpha = 0.97$ [95% CI, 0.97, 0.98]).

Convergent validity

The simulation-derived tool metrics were correlated with the average OSATS ratings to assess convergent validity. A two-tailed Spearman rank order correlation coefficient was calculated between all collected data for each tool metric that achieved evidence of construct validity and each OSATS item.

Statistical analysis

Collected data were imported into Python to develop tool metrics. Outliers in tool metrics were identified and imputed in MATLAB R2023b. All other statistical assessments were performed on SPSS (version 29.0; IBM, Armonk, New York). The data were not normally distributed as assessed by Shapiro-Wilk's test ($P < 0.05$).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:23:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cjn.2025.10404

Table 1. Steps and tools utilized for each pedicle screw insertion simulation employing the TSYM simulator platform

Steps	Objective	Tool required
Step 1: Entry point creation	Choose entry point for the pedicle screw, and verification using fluoroscopy	Awl
Step 2A: Channel Creation	Create channel in the pedicle and verification using fluoroscopy	Pedicle finder
Step 2B: Channel Breach Verification	Check for presence or absence of a pedicle breach	2 mm ball tip probe
Step 3A: Tap Insertion	Pre-thread the previously created channel in the pedicle and verification using fluoroscopy	5.5 mm tap
Step 3B: Pedicle Breach Verification	Check for presence or absence of a pedicle breach	2 mm ball tip probe
Step 4: Screw insertion	Insertion of the selected screw by rotation the screwdriver and verify using fluoroscopy	Screwdriver and Screw (6.5 mm diameter and 4.5 mm length)

Mann–Whitney U-tests assessed statistical differences between groups for each performance measure, and effect sizes for significant findings were reported (Cohen's r). A two-tailed Spearman rank order correlation coefficient examined associations between performance metrics.

Results

Participants

Participants' demographic data and relevant experience are presented in Table 2. A total of 27 participants from two Quebec universities were included. While the participant pool is small, other studies have successfully assessed face, content and construct validity of two different spine surgery virtual reality simulators with a similar participant size.[26,32] The skilled group reported a mean of 452 pedicle screws (SD = 883.6) inserted independently, while the less skilled group reported a mean of 0.5 pedicle screws (SD = 1.4) inserted. The difference between the two groups was statistically significant ($P < 0.001$). Since each participant inserted 4 screws, a total of 108 simulated screws were inserted. One screw was removed from the study due to a technical issue, resulting in 107 screws available for analysis. Therefore, 107 videos, one for each pedicle screw insertion, were evaluated using OSATS.

Face and content validity

The pedicle screw insertion simulation median ratings and ranges for face and content validity are outlined in Table 3. The four participating spine surgeons and two spine fellows assessed face and content validity. This group rated the simulated procedure's overall realism with a 5.0 median (range 3.0–6.0) rating, consistent with face validity. All steps achieved evidence of content validity (median ≥ 4.0) except the pre-threading step using the tap, which was rated a median of 3.5 (range 1.0–5.0). The skilled group rated the simulated procedure's overall realism with a 5.0 median (3.0–6.0) rating.

Construct validity

Simulation-derived tool metrics

All simulation-derived tool metrics were assessed between the groups (Table 4). Significant differences were found between the two groups in 4 of 25 performance metrics. We anticipated

Table 2. Demographic data for the two groups performing the simulated pedicle screw insertion on the TSYM simulator platform

	Less Skilled	Skilled
Number of participants	14 (52%)	13 (48%)
Age (years)		
Mean (SD)	29 (1.7)	38 (8.1)
Gender		
Male	12 (86%)	13 (100%)
Female	2 (14%)	0 (0%)
Specialty		
Neurosurgery	10 (71%)	8 (62%)
PGY 1-4	10	–
PGY 5-6	–	5
Non-spine Fellow	–	2
Spine Surgeon	–	1
Orthopedics	4 (28%)	5 (38%)
PGY 1-4	4	–
PGY 5	–	–
Spine Fellow	–	1
Spine Surgeon	–	4
Affiliation		
McGill	11 (41%)	9 (33%)
Université de Montréal	3 (11%)	4 (15%)
Number of Reported Pedicle Screws Inserted**		
Mean (SD)	0.5 (1.4)	452 (883.6)
Median (Range)	0 (0-5)	100 (10-3000)
Prior Experience with any Virtual Reality Surgical Simulator		
Yes	3 (21%)	5 (38%)
No	11 (79%)	8 (62%)

PGY = Postgraduate year; SD = standard deviation; **No significant difference was found between the two groups except for the mean number of reported pedicle screws inserted ($P < 0.001$).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:23:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cjn.2025.10404

Table 3. Face and content validity

Validity type	Validity statements	Median response of spine fellows and spine surgeons group	Observed range
Content Validity	Using the awl to create the entry point for the pedicle screw.	5.00	(2.0–6.0)
	Using the curved pedicle finder to develop the screw channel in the pedicle.	4.00	(1.0–5.0)
	Using the ball tip probe to assess for pedicle breach in the created channel in the pedicle.	4.00	(2.0–6.0)
	Using the tap to create threads to the inner canal.	3.50	(1.0–5.0)
	Inserting the screw into the created channel in the pedicle.	4.50	(1.0–6.0)
Face Validity	Please rate the overall anatomical realism of the simulated spine.	4.00	(3.0–5.0)
	Please rate the overall realism of the color for the simulated anatomical structures.	4.00	(4.0–6.0)
	Please rate the overall realism of the procedure.	5.00	(3.0–5.0)
	If this simulator was available in your program, you would use this simulation scenario for training of the technical skills simulated.	4.50	(1.0–7.0)

The median score on a 7-point Likert scale for face and content validity for the spine fellows and surgeons after completing the pedicle screw simulation.

observing group differences between 3D velocity and 3D acceleration of the tap screw at step 3A and tool contact and maximum force of the screwdriver in step 4.[33–35] While pre-threading the channel with the tap, the skilled group showed a significant increase in 3D velocity when compared to the less skilled group (0.0014, 95% CI [0.00119, 0.00153] vs 0.001, 95% CI [0.0012, 0.0013]; Cohen's $r = 0.20$; $P = 0.04$). Using the tap, the less skilled group showed a significantly higher 3D acceleration than the skilled group (4.36e-9, 95% CI [-7.26e-9, 16e-9] vs 5.43e-10, 95% CI [-5.19e-9, 6.28e-9]; Cohen's $r = 0.24$; $P = 0.01$). Although the 3D acceleration values were small across both groups, statistical analysis confirmed a significant difference ($P = 0.01$). During the insertion of the screw with the screwdriver, the less skilled group applied significantly more maximum force than the skilled group (10.14, 95% CI [7.34, 12.96] vs 7.52, 95% CI [5.07, 9.96]; Cohen's $r = 0.20$; $P = 0.04$) and spent significantly more time in contact with surrounding tissue than the skilled group (0.22, 95% CI [0.18, 0.25] vs 0.11, 95% CI [0.09, 0.13]; Cohen's $r = 0.47$; $P < 0.001$). These differences are depicted in Figure 2.

Randomized, blinded OSATS ratings

An average rating for each OSATS item was calculated for each screw video by blinded ratings provided by two experts. The skilled group achieved a significantly higher mean overall OSATS rating compared to the less skilled group (5.02, 95% CI [4.63, 5.41] vs 3.30, 95% CI [2.92, 3.69]; $P < .001$). In each OSATS item (instrument handling, respect for tissue, economy of movement, flow and knowledge of procedure), the skilled group significantly outperformed the less skilled group ($P < 0.001$ for each item; respective Cohen's $r = 0.55, 0.43, 0.55, 0.54, 0.52, 0.52$). Group differences are outlined in Figure 3.

Convergent validity

A two-tailed Spearman rank order correlation coefficient was calculated between each item of the OSATS ratings and the four significant tool metrics (screwdriver maximum force, screwdriver tool contact, 3D velocity using the tap and 3D acceleration using the tap). As predicted, the maximum force using the screwdriver had significant negative correlations with all OSATS items: respect for tissue, instrument handling, economy of movement, flow, knowledge of procedure and overall (Spearman's coefficient = −0.32, $P < 0.01$; Spearman's coefficient = −0.39, $P < 0.01$; Spearman's coefficient = −0.37, $P < 0.01$; Spearman's coefficient = −0.38, $P < 0.01$; Spearman's coefficient = −0.29, $P < 0.01$; Spearman's coefficient = −0.33, $P < 0.01$, respectively). As predicted, tool contact using the screwdriver significantly correlated with respect for tissue, instrument handling, economy of movement, flow, knowledge of procedure and overall (Spearman's coefficient = −0.25, $P < 0.01$; Spearman's coefficient = −0.34, $P < 0.01$; Spearman's coefficient = −0.42, $P < 0.01$; Spearman's coefficient = −0.43, $P < 0.01$; Spearman's coefficient = −0.31, $P < 0.01$; Spearman's coefficient = −0.31, $P < 0.01$, respectively). The tap's 3D velocity significantly correlated with four out of six OSATS items, including economy of movement, flow, knowledge of procedure and overall (Spearman's coefficient = 0.29, $P < 0.01$; Spearman's coefficient = 0.25, $P = 0.01$; Spearman's coefficient = 0.21, $P = 0.03$; Spearman's coefficient = 0.20, $P = 0.04$). No significant correlations were found between the 3D acceleration and OSATS items. Table 5 outlines the associations between these performance metrics.

Discussion

The present study offers insight for surgical educators and researchers interested in spine simulation. First, the study's pedicle screw insertion simulation demonstrated varying degrees of validity. Second, to our knowledge, this is the first study to correlate simulator-derived metrics with OSATS ratings to assess the convergent validity in a virtual reality spine platform. Finally, the dual performance assessment approach, using OSATS ratings and simulator-derived metrics, offers a comprehensive understanding of learner-operative performance.

Face, content and construct validity

This study used traditional (face, content and construct validity) and contemporary frameworks to construct a validity argument for the TSYM simulator's use in surgical training.[22–24] Face validity was included as subjective feedback but was not central to the validity argument. OSATS findings provided the strongest support, while evidence from the other measures was less robust, given their variability and small effect sizes.

Face and content validity were supported, with eight of nine statements rated with a median of 4.0 or greater by six participating spine surgeons and fellows.[26,31] However, the variability of the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:23:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cjn.2025.10404

Table 4. Simulation-derived metrics obtained from the L4-L5 bilateral pedicle screw insertion simulation on the TSYM simulator and corresponding Mann–Whitney U-test P-value

Tool and Metrics	*P*-value
Awl	
3D Velocity	0.75
3D Force	0.23
Max Force	0.37
3D Acceleration	0.16
Tool Contact	0.51
Pedicle finder	
3D Velocity	0.71
3D Force	0.12
Max Force	0.54
3D Acceleration	0.52
Tool Contact	0.28
Ball Tip Probe	
3D Velocity	0.10
3D Force	0.12
Max Force	0.92
3D Acceleration	0.23
Tool Contact	0.31
Tap Screw	
3D Velocity	0.04*
3D Force	0.40
Max Force	0.37
3D Acceleration	0.01*
Tool Contact	0.45
Screwdriver	
3D Velocity	0.52
3D Force	0.12
Max Force	0.04*
3D Acceleration	0.94
Tool Contact	<0.001*

*Significant *p*-value for Mann–Whitney *U*-test, nonparametric test ($P < 0.05$).

results was wide (range: 1–7), and expert verbal feedback indicated that torque feedback from the tap for pre-threading the inner pedicle canal could be improved. Thus, the present results must be interpreted with care.

For construct validity, 4 of 25 simulation-derived tool metrics significantly distinguished the two groups with small effect sizes: 3D velocity and 3D acceleration of the simulated tap screw, and the maximum force and the tool contact of the simulated screwdriver. The skilled group exhibited higher 3D velocity and lower acceleration with tap screw use than the less skilled. These patterns are associated with previous studies showing smoother, controlled movements among surgical experts.[8,9,33] Conversely, the less skilled group's unfamiliarity with this instrument may have resulted in lower tap velocity. Meanwhile, the maximum force applied by the screwdriver was significantly higher for the less skilled group than for the skilled group. This is consistent with previous virtual reality studies,[33–37] which show that more skilled participants tend to apply less instrument force, recognizing that excessive force may compromise patient safety.[35] The less skilled group's higher screwdriver contact could likely be attributed to less precision, causing unintended tissue contact. The skilled group significantly outperformed the less skilled group in each OSATS component (Figure 3). These findings provide evidence of construct validity for the TSYM simulator's pedicle screw insertion simulation.

Correlating simulation-derived performance metrics and OSATS ratings for convergent validity

Three of four simulation-derived performance metrics significantly correlated with all OSATS items, with moderate effect sizes, providing evidence of convergent validity for the TSYM simulator and suggesting several important implications. Screwdriver maximum force and tool contact were negatively correlated with all OSATS items, while 3D velocity using the tap positively correlated with four OSATS items: the economy of movement, flow, knowledge of procedure and overall score, supporting convergent validity. The less skilled groups' lower OSATS ratings were consistent with their poorer performance on these key simulation-derived metrics. Instrument handling and respect for tissue did not significantly correlate with the 3D velocity using the tap, while its 3D acceleration did not significantly correlate with any OSATS item. These findings suggest that OSATS may not fully capture key performance features, possibly due to limitations of visual assessment in evaluating instrument dynamics, like acceleration within the bone channel.[35,38] Although OSATS is a validated tool for assessing surgical performance, several studies have questioned its ability to reflect the full complexity of operative performance.[39,40] This study indicates that combining OSATS with simulator-derived metrics could provide a more formative and comprehensive approach to evaluating and improving surgical skills. It also provides support for further research on the TSYM simulator's potential to predict future pedicle screw insertion performance in patients.

TSYM as an educational tool

The results suggest that the TSYM simulator pedicle screw insertion scenario may be useful for the evaluation and training of less skilled learners, specifically on the four metrics showing construct validity. Virtual reality simulators have been assessed in pedicle screw placement training and have improved the accuracy and skill acquisition of pedicle screw placement.[5,6,21,41,42] Further, incorporating virtual reality simulation into the spine surgery learning curriculum may benefit less skilled trainees by providing a valuable platform for practicing complex spine procedures and supporting formative skill development.[20,21] However, the TYSM simulator pedicle screw insertion scenario may benefit from modification to meet its full potential as a surgical education system.

This study's findings align with prior research on neurosurgical simulators. A systematic review found that while the visual appearance of neurosurgical virtual reality simulators is generally favorable, haptic feedback remains a limitation across simulation platforms.[43] This aligns with this study's face and content validity results, where haptic-related features showed a greater variability in expert responses. Related to construct and convergent validity, Ledwos et al. demonstrated that skilled participants utilized greater maximum force than less skilled participants on a virtual reality spine simulator, while a systematic review identified maximum

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:23:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cjn.2025.10404

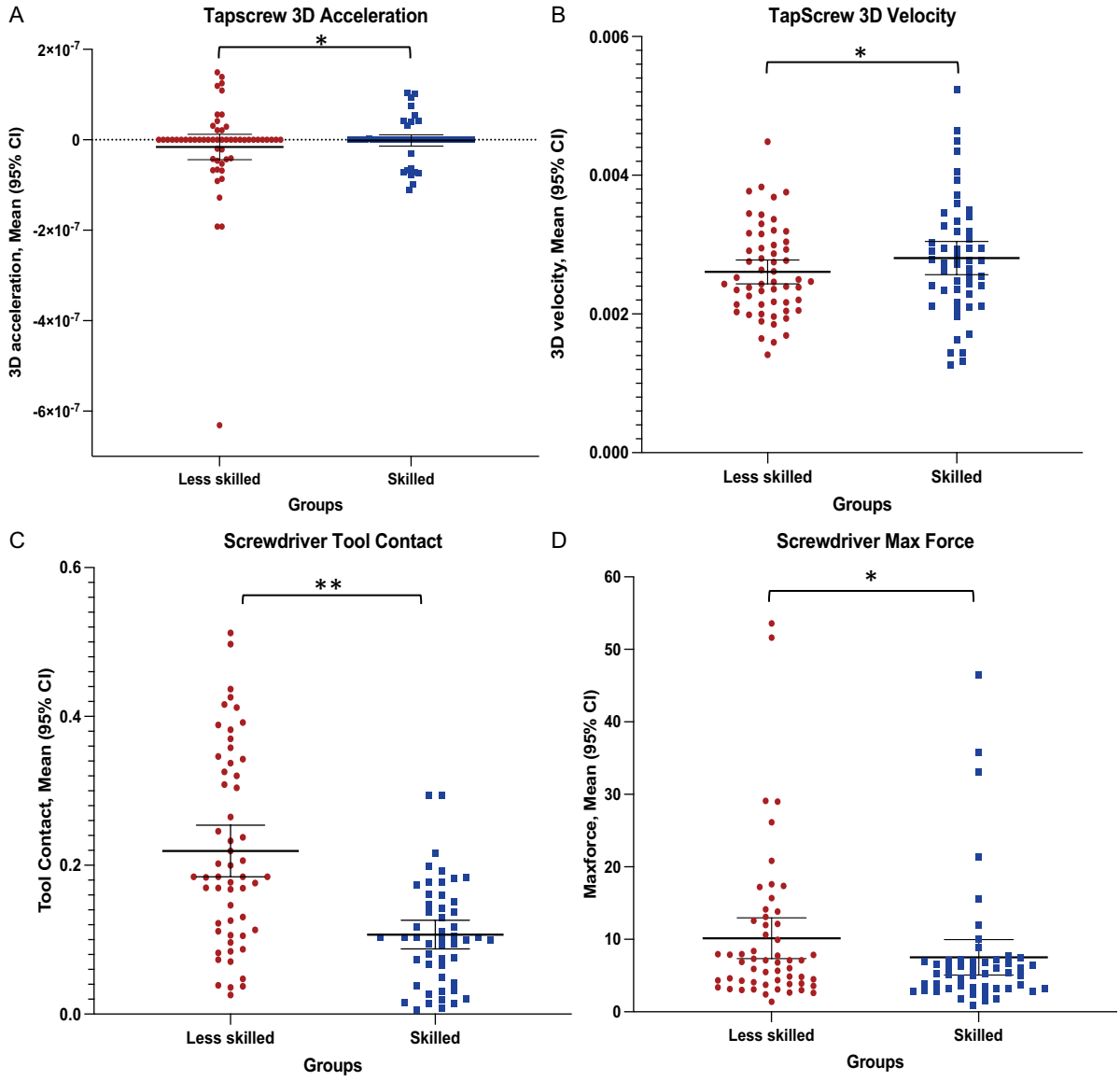


Figure 2. Significant performance assessments of the task using simulation-generated performance metrics. (A) Tap screw's 3D velocity. (B) Tap screw's 3D acceleration. (C) Screwdriver max force on the pedicle. (D) Screwdriver contact with pedicle. The central line indicates the mean value for each group. *Represents a significant difference between groups after Mann–Whitney U, nonparametric test ($P < 0.05$). **Represents a significant difference between groups after Mann–Whitney U, nonparametric test ($P < 0.01$).

force as a reliable indicator of surgical expertise.[26,44] Further, an umbrella review suggests that performance metrics related to force and kinematics effectively ascertain skill level.[45] Another systematic review found that neurosurgical virtual reality simulators' performance metrics correlate well with intraoperative skills.[43] Together, these studies support our investigation's convergent validity findings, showing that key simulator-derived metrics, particularly those related to force and motion, align with OSATS ratings and can effectively distinguish between levels of expertise.

With the vast data generated from virtual reality simulators like the TSYM platform, artificial intelligence (AI) methodologies may enhance the understanding of surgical skills' precision and granularity.[35,36,41,46] Further, it can be utilized to create intelligent tutoring systems, like the Intelligent Continuous Expertise Monitoring System.[42] However, incorporating human educator input is essential, as these systems have been linked to unintended outcomes.[47,48] A recent randomized clinical trial demonstrated that AI-augmented personalized expert instruction resulted in improved simulated surgical performance, suggesting that spine simulation platforms may benefit from utilizing these technologies in future studies and curriculum design.[49] Deep learning models that integrate simulator-derived metrics and equivalent OSATS video ratings may enable future AI systems to predict OSATS scores only using simulator data.[48] Finally, implementing this data with intelligent tutoring systems can contribute to developing an "Intelligent Operating Room" that continually assesses and trains learners while minimizing surgical errors.[31,38,41,50]

Limitations

The TSYM simulation platform has limitations. The pedicle screw insertion simulation does not capture the dynamic intraoperative learning environment, the flexible sequence during human

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:23:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cjn.2025.10404

Table 5. Convergent validity determination between simulation-derived performance metrics and OSATS scoring

Simulation derived performance metrics[a]	OSATS Scoring											
	Respect for tissue		Instrument handling		Economy of movement		Flow		Knowledge of Procedure		Overall	
	Spearman's Coefficient	ρ Value	Spearman's Coefficient	ρ Value	Spearman's Coefficient	ρ Value	Spearman's Coefficient	ρ Value	Spearman's Coefficient	ρ Value	Spearman's Coefficient	ρ Value
Screwdriver Maximum Force	−0.32	<0.01**	−0.39	<0.01**	−0.37	<0.01**	−0.38	<0.01**	−0.293	<0.01**	−0.33	<0.01**
Screwdriver Tool Contact	−0.25	0.01*	−0.34	<0.01**	−0.42	<0.01**	−0.43	<0.01**	−0.31	<0.01**	−0.31	<0.01**
Tap 3D Velocity	0.18	0.06	0.10	0.29	0.29	<0.01**	0.25	0.01*	0.21	0.03*	0.21	0.04*
Tap 3D Acceleration	−0.14	0.16	−0.13	0.18	−0.17	0.09	−0.15	0.13	−0.15	0.13	−0.15	0.14

*Significant ρ-value for Spearman's rank coefficient of correlation ($\rho < 0.05$). ** Significant ρ-value for Spearman's rank coefficient of correlation ($\rho < 0.01$). [a]Simulation-derived performance metrics that showed construct validity. OSATS = Objective Structured Assessment of Technical Skills.

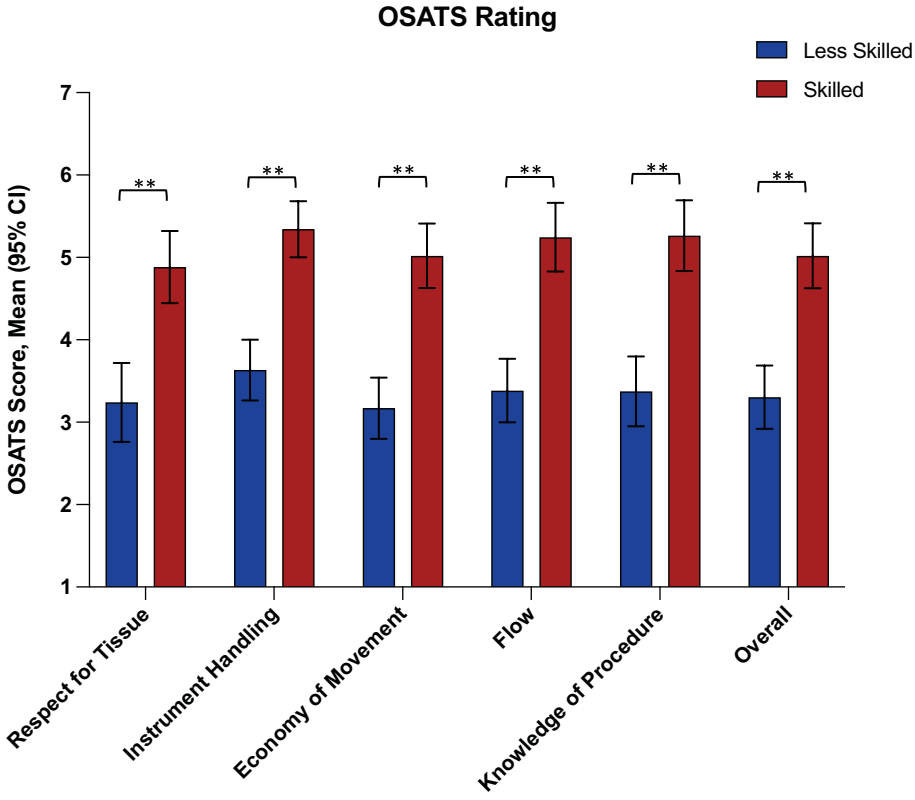


Figure 3. Performance assessment of the pedicle screw insertion task using OSATS. *Represents a significant difference between groups after Mann–Whitney U-test, nonparametric test ($P < 0.05$). **Represents a significant difference between groups after Mann–Whitney U-test, nonparametric test ($P < 0.01$). OSATS = Objective Structured Assessment of Technical Skills.

procedures and bimanual psychomotor skills utilized during patient spinal procedures, given its single-handed robotic arm setup.[20,38,40] The present study's sample size was limited due to clinical commitments, limiting the generalization of results. Further, the statistical analyses for construct and convergent validity may have been underpowered, with significant findings possibly due to Type I error and reflected in the low to moderate effect sizes. While common in surgical education research, this limitation underscores the need for larger, multi-institutional samples to improve robustness and generalizability.[46] Additionally, the study may be subject to potential biases, such as preconceived notions and social desirability bias, as face and content validity were measured through self-reports.[51,52] In this study, each pedicle screw insertion was evaluated individually due to variations in entry points, screw angulation and anatomy. Larger studies are needed to evaluate how repeated insertions affect the learning curves of less skilled and skilled individuals. Finally, to standardize the procedure, participants used a fixed-size screw, despite the TSYM platform offering various screw sizes and lengths to assess procedural skill.

Conclusion

While several limitations and challenges exist with the TSYM simulator platform pedicle screw insertion scenario, some

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:23:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cjn.2025.10404

performance metrics, including screwdriver maximum force, screwdriver tool contact and Tap 3D velocity, show potential to assist in surgical teaching. The information garnered from this study may allow improvements in the TSYM simulator to optimize future performance.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/cjn.2025.10404.

Acknowledgments. The authors would like to thank all the neurosurgeons and orthopedic spine surgeons, fellows, along with neurosurgical and orthopedic residents who participated in this study. Special thanks to Dr Ahmed Aoude for allowing the authors to use the Orthopedic Research Laboratory, Montreal General Hospital, for these studies, and Dr Greg Berry for helping organize the use of orthopedic facilities and recruiting orthopedic residents for the study. The authors also thank Dr Zhi Wang, Dr Sung-Joo Yuh, Dr Ahmed Aoude, Dr Lucy Luo, Dr Ahmad Alsayegh, Dr Mohamad Bakhaidan, Dr Carlo Santaguida and Dr Abdulrahman Almansouri for their help recruiting trial participants. Finally, a special thanks to Dr Jason Harley for their expert educational input and to Dr Jose Correa for providing statistical input. This study was supported by Mitacs Accelerate Grant, Brain Tumour Foundation of Canada-Brain Tumour Research Grant, a Medical Education Research Grant from the Royal College of Physicians, the Franco Di Giovanni Foundation and the Montreal Neurological Institute and Hospital, McGill University. Cedarome Canada Inc. dba Symgery supplied the TSYM Symgery virtual reality nonimmersive simulator platform utilized for these investigations.

Author contributions. Trisha Tee: Contributed to conceptualization, methodology, data collection, formal analysis, investigation and writing. Noel Abboud: Contributed to methodology, formal analysis and writing. Bilal Tarabay: Contributed to conceptualization and methodology, formal analysis, data collection, participant recruitment and writing – review and editing. Abdulmajeed Albeloushi: Contributed to conceptualization and methodology, data collection and participant recruitment. Puja Pachchigar: Contributed to conceptualization and methodology, formal analysis, data collection and processing and participant recruitment. Mohamed Alhantoobi: Contributed to conceptualization and methodology, and formal analysis. Nour Abou Hamdan: Contributed to conceptualization and methodology and formal analysis. Recai Yilmaz: Contributed to conceptualization and methodology, formal analysis, and writing – review and editing. Ali Fazlollahi: Contributed to conceptualization and methodology. Rolando F. Del Maestro: Contributed to project creation, conceptualization, methodology, resources, investigation, project funding, guidance, supervision of this research, interpreting results, writing – original draft and writing – review and editing.

Funding statement. Trisha Tee, Bilal Tarabay and Puja Pachchigar are supported by a Mitacs Accelerate Internship Grant. Trisha Tee also received support from a Masters-CIHR. Dr Recai Yilmaz was supported by a Brain Tumour Foundation of Canada-Brain Tumour Research Grant, a Medical Education Research Grant from the Royal College of Physicians, a Max Binz Fellowship from McGill University Internal Studentships and a grant from the Fonds de recherche du Quebec–Santé. Dr Rolando Del Maestro is affiliated with a CIHR-funded grant but did not receive direct financial support from this grant. The Franco Di Giovanni Foundation supported the lab computer technology, and the Montreal Neurological Institute and Hospital provided lab space. The authors have no personal, financial or institutional interest in any of the drugs, materials or devices described in this article.

Competing interests. Dr Rolando Del Maestro and Dr Recai Yilmaz are co-inventors for pending patents related to training platforms and intelligent monitoring systems with patent numbers 05001770-843USPR and 05001770-883USPR, respectively. These patents are not associated with the study. Dr Rolando Del Maestro held positions as the President American Osler Society from 2023–2024 and is currently a Member Board of the American Osler Society since 2022, Chairperson Osler Library Standing Committee since 2015, and a Member Board of the Osler Library at McGill University since 2015. No payments were received from these roles.

References

1. Leung A, Luu S, Regehr G, Murnaghan ML, Gallinger S, Moulton C-A. First, do no harm": balancing competing priorities in surgical practice. *Acad Med*. 2012;87:1368–1374.
2. Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. *Acad Med*. 2003;78:783–788.
3. Rattner DW, Apelgren KN, Eubanks WS. The need for training opportunities in advanced laparoscopic surgery. *Surg Endosc*. 2001;15:1066–1070.
4. Daniels AH, Ames CP, Garfin SR, et al. Spine surgery training: is it time to consider categorical spine surgery residency? *The Spine J*. 2015;15: 1513–1518.
5. Grantcharov TP, Reznick RK. Teaching procedural skills. *BMJ*. 2008;336:1129–1131.
6. McGaghie WC. Mastery learning: it is time for medical education to join the 21st century. *Acad Med*. 2015;90:1438–1441.
7. Manbachi A, Cobbold RS, Ginsberg HJ. Guided pedicle screw insertion: techniques and training. *Spine J*. 2014;14:165–179.
8. Gang C, Haibo L, Fancai L, Weishan C, Qixin C. Learning curve of thoracic pedicle screw placement using the free-hand technique in scoliosis: how many screws needed for an apprentice? *Eur Spine J*. 2012;21:1151–1156.
9. Gonzalvo A, Fitt G, Liew S, et al. The learning curve of pedicle screw placement: how many screws are enough? *Spine*. 2009;34:E761–E765.
10. Gautschi OP, Schatlo B, Schaller K, Tessitore E. Clinically relevant complications related to pedicle screw placement in thoracolumbar surgery and their management: a literature review of 35,630 pedicle screws. *Neurosurg Focus*. 2011;31:E8.
11. Hicks JM, Singla A, Shen FH, Arlet V. Complications of pedicle screw fixation in scoliosis surgery: a systematic review. *Spine (Phila Pa 1976)*. 2010;35:E465–E470.
12. Alotaibi FE, AlZhrani GA, Sabbagh AJ, Azarnoush H, Winkler-Schwartz A, Del Maestro RF. Neurosurgical assessment of metrics including judgment and dexterity using the virtual reality simulator neuroTouch (NAJD metrics). *Surg Innov*. 2015;22:636–642.
13. Rogers MP, DeSantis AJ, Janjua H, Barry TM, Kuo PC. The future surgical training paradigm: virtual reality and machine learning in surgical education. *Surgery*. 2021;169:1250–1252.
14. Jung Y, Muddaluru V, Gandhi P, Pahuta M, Guha D. The development and applications of augmented and virtual reality technology in spine surgery training: a systematic review. *Can J Neurol Sci/Journal Canadien des Sciences Neurologiques*. 2024;51:255–264
15. Yuk FJ, Maragkos GA, Sato K, Steinberger J. Current innovation in virtual and augmented reality in spine surgery. *Ann Transl Med*. 2021;9:94.
16. Pfandler M, Lazarovici M, Stefan P, Wucherer P, Weigl M. Virtual reality-based simulators for spine surgery: a systematic review. *Spine J*. 2017;17:1352–1363.
17. Wang Z, Shen J. Simulation training in spine surgery. *J Am Acad Orthop Surg*. 2022;30:400–408.
18. Asoodar M, Janesarvatan F, Yu H, de Jong N. Theoretical foundations and implications of augmented reality, virtual reality, and mixed reality for immersive learning in health professions education. *Adv Simul (Lond)*. 2024;9:36.
19. Sewell C, Morris D, Blevins NH, et al. Providing metrics and performance feedback in a surgical simulator. *Comput Aided Surg*. 2008;13:63–81.
20. AlOtaibi F, Al Zhrani G, Bajunaid K, Winkler-Schwartz A, Azarnoush H, Mullah M. Assessing neurosurgical psychomotor performance: role of virtual reality simulators, current and future potential. *SOJ Neurol*. 2015;2:1–7.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:23:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cjn.2025.10404

21. Azarnoush H, Alzhrani G, Winkler-Schwartz A, et al. Neurosurgical virtual reality simulation metrics to assess psychomotor skills during brain tumor resection. *Int J Comput Assist Radiol Surg*. 2015;10:603–618.
22. Huang C, Cheng H, Bureau Y, Ladak HM, Agrawal SK. Automated metrics in a virtual-reality myringotomy simulator: development and construct validity. *Otol Neurotol*. 2018;39:e601–e608.
23. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc Interv Tech*. 2003;17:1525–1529.
24. Fried MP, Sadoughi B, Weghorst SJ, et al. Construct Validity of the Endoscopic Sinus Surgery Simulator: II. Assessment of Discriminant Validity and Expert Benchmarking. *Arch Otolaryngol Head Neck Surg*. 2007;133:350–357.
25. Schout BM, Hendrikx AJ, Scheele F, Bemelmans BL, Scherpbier AJ. Validation and implementation of surgical simulators: a critical review of present, past, and future. *Surg Endosc*. 2010;24:536–546.
26. Ledwos N, Mirchi N, Bissonnette V, Winkler-Schwartz A, Yilmaz R, Del Maestro RF. Virtual reality anterior cervical discectomy and fusion simulation on the novel sim-ortho platform: validation studies. *Oper Neurosurg (Hagerstown)*. 2020;20:74–82.
27. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med*. 1996;71:1363–1365.
28. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg*. 2010;97:972–987.
29. Orovec A, Bishop A, Scott SA, et al. Validation of a surgical objective structured clinical examination (S-OSCE) using convergent, divergent, and trainee-based assessments of fidelity. *J Surg Educ*. 2022;79:1000–1008.
30. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61:344–349.
31. Almansouri A, Abou Hamdan N, Yilmaz R, et al. Continuous instrument tracking in a cerebral corticectomy ex vivo calf brain simulation model: face and content validation. *Oper Neurosurg (Hagerstown)*. 2024;27:106–113.
32. Alkadri S, Del Maestro RF, Driscoll M. Face, content, and construct validity of a novel VR/AR surgical simulator of a minimally invasive spine operation. *Med Biol Eng Comput*. 2024;62:1887–1897.
33. Ebina K, Abe T, Higuchi M, et al. Correction to: motion analysis for better understanding of psychomotor skills in laparoscopy: objective assessment-based simulation training using animal organs. *Surg Endosc*. 2021;35:4417.
34. Sawaya R, Bugdadi A, Azarnoush H, et al. Virtual reality tumor resection: the force pyramid approach. *Oper Neurosurg (Hagerstown)*. 2018;14:686–696.
35. Mirchi N, Bissonnette V, Ledwos N, et al. Artificial neural networks to assess virtual reality anterior cervical discectomy performance. *Oper Neurosurg*. 2020;19:65–75.
36. Alkadri S, Ledwos N, Mirchi N, et al. Utilizing a multilayer perceptron artificial neural network to assess a virtual reality surgical procedure. *Comput Biol Med*. 2021;136:104770.
37. Reich A, Mirchi N, Yilmaz R, et al. Artificial neural network approach to competency-based training using a virtual reality neurosurgical simulation. *Oper Neurosurg*. 2022;23:31–39.
38. Mirchi N, Bissonnette V, Yilmaz R, Ledwos N, Winkler-Schwartz A, Del Maestro RF. The virtual operative assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. *PLoS ONE*. 2020;15:e0229596.
39. Bernard JA, Dattilo JR, Srikumaran U, Zikria BA, Jain A, LaPorte DM. Reliability and validity of 3 methods of assessing orthopedic resident skill in shoulder surgery. *J Surg Educ*. 2016;73:1020–1025.
40. Anderson DD, Long S, Thomas GW, Putnam MD, Bechtold JE, Karam MD. Objective structured assessments of technical skills (OSATS) does not assess the quality of the surgical result effectively. *Clin Orthop Relat Res®*. 2016;474:874–881.
41. Yilmaz R, Winkler-Schwartz A, Mirchi N, et al. Continuous monitoring of surgical bimanual expertise using deep neural networks in virtual reality simulation. *npj Digital Medicine*. 2022;5:54.
42. Hou Y, Lin Y, Shi J, Chen H, Yuan W. Effectiveness of the thoracic pedicle screw placement using the virtual surgical training system: a cadaver study. *Oper Neurosurg (Hagerstown)*. 2018;15:677–685.
43. Chawla S, Devi S, Calvachi P, Gormley WB, Rueda-Esteban R. Evaluation of simulation models in neurosurgical training according to face, content, and construct validity: a systematic review. *Acta Neurochir (Wien)*. 2022;164:947–966.
44. Chan J, Pangal DJ, Cardinal T, et al. A systematic review of virtual reality for the assessment of technical skills in neurosurgery. *Neurosurg Focus*. 2021;51:E15.
45. Harley JM, Tawakol T, Azher S, Quaiattini A, Del Maestro R. The role of artificial intelligence, performance metrics, and virtual reality in neurosurgical education: an umbrella review. *Glob Surg Edu - J Assoc Surg Edu*. 2024;3:83.
46. Winkler-Schwartz A, Yilmaz R, Mirchi N, et al. Machine learning identification of surgical and operative factors associated with surgical expertise in virtual reality simulation. *JAMA Network Open*. 2019;2:e198363.
47. Fazlollahi AM, Yilmaz R, Winkler-Schwartz A, et al. AI in surgical curriculum design and unintended outcomes for technical competencies in simulation training. *JAMA Network Open*. 2023;6:e2334658.
48. Yilmaz R, Bakhaidar M, Alsayegh A, et al. Real-time multifaceted artificial intelligence vs in-person instruction in teaching surgical technical skills: a randomized controlled trial. *Sci Rep*. 2024;14:15130.
49. Giglio B, Albeloushi A, Alhaj AK, et al. Artificial intelligence–augmented human instruction and surgical simulation performance: a randomized clinical trial. *JAMA Surg*. 2025 [Epub ahead of print 2025 Aug 6]. doi:10.1001/jamasurg.2025.2564.
50. Fazlollahi AM, Bakhaidar M, Alsayegh A, et al. Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: a randomized clinical trial. *JAMA Network Open*. 2022;5:e2149008.
51. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol*. 1998;2:175–220.
52. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. *J Multidiscip Healthc*. 2016;9:211–217.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 15:23:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cjn.2025.10404