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

Trisha Tee, BSc

Department of Experimental Surgery

Faculty of Medicine and Health Sciences

McGill University, Montreal

August 2024

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Experimental Surgery

Table of Contents

ABSTRACT

Objective: Virtual reality spine simulators have the potential to become valuable educational tools, offering learners a safe, risk-free environment to assess and train their psychomotor skills in challenging operative procedures like pedicle screw insertions. The TSYM Symgery simulator platform is a virtual reality spine simulator capable of deconstructing and simulating complex spine procedures, including pedicle screw insertions. This case series study aims to investigate the face, content, construct, and convergent validity of an L4-L5 bilateral pedicle screw insertion on the TSYM simulator platform.

Methods: Neurosurgical and 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 groups (post-graduate year (PGY) 5-6, fellows, and consultant neurosurgeons or orthopedic surgeons) or less skilled (PGY 1-4). Face and content validity were assessed utilizing a 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. The skilled group rated all face and content validity statements with a median ≥ 4. Significant differences between the less skilled and skilled groups were found for 4 of 25 simulation-derived performance metrics ($P < .05$) and all OSATS categories ($P < .001$). Two simulation-derived performance metrics (maximum force and tool contact using the simulated screwdriver) significantly correlated with OSATS ratings consistent with convergent validation.

Conclusion: The L4-L5 bilateral pedicle screw insertion simulation on the TSYM Symgery simulation platform demonstrated mixed and variable evidence for face, content, construct, and convergent validity, supporting some degree of educational potential for spine surgery training. Improvements are needed to optimize the potential of the TSYM Symgery simulator platform.

RESUMÉ

Objectif : Les simulateurs de colonne vertébrale en réalité virtuelle ont le potentiel de devenir des outils éducatifs précieux offrant un environnement sûr et sans risque pour évaluer et former les compétences psychomotrices des jeunes chirurgiens dans des procédures opératoires complexe comme les insertions de vis pédiculaires. Le simulateur TSYM Symgery propose une platforme de réalité virtuelle capable de déconstruire et de simuler des procédures complexes en chirurgie rachidienne, y compris les insertions de vis pédiculaires. Cette série de cas vise à examiner la validité de face, contenu, construit et de convergence d'une insertion bilatérale de vis pédiculaires L4-L5 sur la plateforme de simulateur TSYM.

Méthodes : Des résidents en neurochirurgie et en orthopédie, ainsi que des fellows et des chirurgiens rachidiens ont effectué des insertions bilatérales de vis pédiculaires L4-L5 sur le simulateur TSYM. Les participants ont été classés en groupes compétents (résidents en PGY 5-6, fellows en chirurgie rachidien et neurochirurgiens consultants ou chirurgiens orthopédistes) ou moins compétents (résidents en PGY 1-4). La validité de face et contenu ont été évaluée en utilisant une échelle de Likert. La validité de construit a été déterminée en examinant les différences de métriques de performance dérivées de la simulation et l'Évaluation Structurée Objective des Compétences Techniques (OSATS). La validité convergente a été examinée en corrélant les métriques de performance dérivées de la simulation et les évaluations OSATS.

Résultats : Treize participants compétent et 14 moins compétents ont été inclus dans cette étude. Le groupe compétent a évalué toutes les déclarations de validité de face et de contenu avec une médiane ≥4. Des différences significatives entre les groupes moins compétent et compétent ont été trouvées pour 4 des 25 métriques de performance dérivées de la simulation et toutes les catégories OSATS, $P < .05$. Les métriques de performance dérivées de la simulation (accélération 3D et vitesse 3D en utilisant le robinet simulé et force maximale et contact avec l'outil en utilisant le tournevis simulé) ont significativement corrélé avec les évaluations OSATS, cohérentes avec la validation convergente.

Conclusion : La simulation de l'insertion bilatérale de vis pédiculaires L4-L5 sur la plateforme de simulation TSYM Symgery a démontré des preuves de validité de face, de contenu, de construit et convergente, soutenant son potentiel comme outil éducatif formateur dans la formation en chirurgie de la colonne vertebrale.

ACKNOWLEDGMENTS

The successful completion of this thesis would not have been possible without numerous individuals' invaluable support and contribution. First and foremost, I would like to express my deepest gratitude to my supervisor and mentor, Dr. Rolando F. Del Maestro. Since I joined the Neurosurgical Simulation and Artificial Intelligence Learning Centre, you have been nothing short of supportive. Thank you for believing in my capabilities and for always having my best interests in mind. Your guidance has been crucial to my achievements and has enabled me to develop both personally and as a researcher, and I am deeply grateful for that. While our discussions have spanned topics from research to art to medicine, I will carry with me all the lessons you have imparted. I would also like to thank Pam Del Maestro for always caring for and supporting all of us in the lab.

I would not be involved in research if it were not for my previous principal investigator, Dr. Said Kourrich. You believed in my capabilities and granted me research opportunities at a time when opportunities were limited. Thank you for instilling confidence in me, spurring my research interest, and playing a monumental role in my success thus far.

I have had the pleasure of working alongside the kind, innovative, and bright members of the Neurosurgical Simulation and Artificial Intelligence Learning Centre. I would like to first thank my colleague, Dr. Recai Yilmaz. You were the first person who introduced me to the lab, and since then, you have constantly helped me with research and achieving my career goals. I appreciate you for being my unofficial mentor and look forward to seeing all you accomplish. Nour Abou Hamdan, Puja Pachchigar, and Chinyelum Agu, thank you for your support, fun adventures, and exciting discussions in and out of the lab. You have made my experience as a graduate student fulfilling and enriching, and I am immensely grateful for each of your friendships. Dr. Bilal Tarabay, Ali Fazlollahi, Dr. Abdulrahman Almansouri, Dr. Abdulmajeed Albeloushi, and Dr. Mohamed Alhantoobi, I appreciate each of you for your valuable insights, intriguing stories, and interesting discussions that have not only helped me as a researcher but have also made the lab environment even more enjoyable. To Vanja Davidovic, Neevya Balasubramaniam, and Bianca Giglio, thank you for your unwavering encouragement and support this past year. I look forward to celebrating each of your achievements. Noel Abboud, it was a pleasure working with you. Thank you for being a key contributor to my thesis, an exceptional collaborator, and great company.

My sincerest thanks to Mitacs for a Mitacs Accelerate Internship Grant for funding this research, enabling me to contribute to the surgical education research field while developing critical skills that I will carry forward. I would also like to express my gratitude to Symgery for collaborating with my team and graciously providing the Neurosurgical Simulation and Artificial Intelligence Learning Centre with your innovative simulator, which was used in the study. I appreciate your commitment to improving surgical training and letting my team be a part of this.

Thank you to the Canadian Institute of Health Research for awarding me the Canada Graduate Scholarship Master's grant. This award has supported my work, and I am honored to be a recipient.

I would also like to thank the orthopedic and neurosurgical residents, fellows, and consultant surgeons who participated in this study, and Dr. Ahmed Aoude and Lucy Luo for playing an instrumental role in recruitment. This study would not be possible without all of you.

My gratitude extends to members of my research advisory committee, Dr. David Labbé, Dr. Gregory Berry, and Dr. Soo-jung Yuh, for their invaluable feedback and commitment to surgical education. I would also like to thank Dr. Jason Harley for supporting me throughout my master's degree and involving me with the SAILS Lab.

Finally, I would like to thank my family and friends for their constant encouragement and support. To my parents, Joy and Howard, and my sisters, Tiffany, Taylor, and Teana, I am beyond grateful to each of you for always supporting my academic endeavors and aspirations. Thank you for your guidance and for consistently inspiring me to pursue my goals. I owe my success to each of you. Emily Rowe and Khalil Virji, I am grateful for your patience, continuous encouragement, and positive attitudes, which have been instrumental in helping me complete my master's degree.

PREFACE AND CONTRIBUTION OF AUTHORS

The structure of this thesis follows a manuscript-based format, and the authors of the manuscript have made substantial contributions to finalizing this work. The author's contributions are detailed using the CRediT (Contributor Roles Taxonomy) format[1,2]. The following statements outline the specific contributions to this research project made by each individual.

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 & editing.

Abudlmajeed Abeloushi: 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 & editing.

Ali Fazlohllahi: Contributed to conceptualization and methodology.

Rolando Del Maestro: Contributed to project creation, conceptualization, methodology, resources, and investigation, project funding, guidance, and supervision of this research, interpreting results, writing - original drafts and writing – review & editing.

ABBREVIATIONS

OSATS: Objective Structured Assessment of Technical Skills

PGY: Post-Graduate Year

3D: Three Dimensional

THESIS INTRODUCTION

Mastering technical skills is an essential learning objective in surgical training, as technical errors can contribute to poor patient outcomes[3,4]. Historically, acquiring technical skills follows an apprenticeship model whereby surgical residents undergo a fixed-length residency learning from a series of educators[5,6]. However, surgical education is transitioning toward a competency-based framework, valuing quantifiable measures of proficiency[7,8].

Tools capable of measuring meaningful performance metrics are a vital component of competency-based training[7]. Virtual reality simulators for technical skill development may be a valuable instrument in this framework[9]. To be implemented in surgical training, virtual reality simulators must undergo a series of validation studies to elucidate their role in surgical curricula[10]. The initial phases of validation involve investigating for face, content, construct, and convergent validity[10]. Establishing these principles forms the groundwork for determining a simulator's educational potential in surgical training[10,11].

In neurosurgery and orthopedic surgery, the pedicle screw insertion is a fundamental technical skill with a steep learning curve[12,13]. Virtual reality simulators may be useful in learning pedicle screw insertions, as it provides a controlled environment to focus on skill development[14]. A limited number of virtual reality simulators for pedicle screw insertions exist, and they lack comprehensive validation studies[15,16,17,18,19]. This limits their ability to be implemented into surgical training[20-22].

The TSYM simulator is a non-immersive, virtual reality platform capable of deconstructing an L4-L5 bilateral pedicle screw insertion. It comprises a single robotic arm that provides haptic feedback during the simulated operation. This new platform has the potential to be a valuable, formative tool in surgical training, specifically for learning pedicle screw insertions, a technically challenging and high-risk technique[9,12,13,10]. However, its potential in surgical training is yet to be explored.

The following study investigates the educational potential of the TSYM simulator's L4-L5 bilateral pedicle screw insertion scenario for neurosurgical and orthopedic residents. This thesis aims to establish the initial validation phases for the TSYM simulator's L4-L5 bilateral pedicle screw insertion, laying the foundation for future studies that can further elucidate its role in surgical education.

BACKGROUND

Surgical Education

Surgical education involves the simultaneous mastery of complex skills, experienced and taught knowledge, and composure in an unpredictable and, at times, highly stressful environment[23]. It is defined as a life-long learning process that begins in residency and continues during the surgeon's career[23]. Since its inception over 100 years ago, its founding principles remain, but its framework has begun to evolve in the last two decades.

The development of the modern surgical residency model can be traced back to the early 1890s by Dr. William Halsted, who at the time was surgeon-in-chief and a Professor of Surgery at Johns Hopkins University[5]. Inspired by the residency program created by his colleague and chief of medicine at Hopkins, Sir William Osler, Dr. William Halsted introduced the Halstedian training model, a pyramidal approach whereby trainees gained increasing responsibility after each training year[5,6]. The principles of this model included acquiring knowledge of surgical disease, skills in patient management, and technical skills with increasing proficiency and independence through repetitive, supervised opportunities to take care of surgical patients[5,6]. Learning under the expert surgeon involved the "see one, do one, teach one" concept, where the surgical trainee is expected to observe a skill, perform the procedure, and be able to consequently teach it[24]. Moreover, Dr. Halsted introduced a structured education with an overarching apprenticeship principle for surgical training, which remains the foundation of surgical education to this day[5,6].

At the present time, surgical residency largely follows the principles it was founded upon. Residents undergo a defined training period at university, university-affiliated, or community hospitals with varying lengths, patient populations, and exposures[3]. Skill and knowledge acquisition are still based on the apprenticeship model, whereby trainees learn under expert surgeons and progressively gain more patient care responsibilities and independence in the operating room[3]. Surgical residency programs also continue to include grand rounds, educational meetings where residents, surgeons, and healthcare providers discuss cases, recent advancements in the field, and relevant research, as a vital component of the curriculum[3]. However, modern surgical training has advanced in its educational framework, including but not limited to protected education time for lectures and journal clubs to enhance critical analysis and appraisal as well as the incorporation of feedback, a critical component for trainee improvement[3]. While this framework has produced many excellent surgeons and favorable outcomes for patients, the current state of surgical education is not without many challenges.

Challenges in Surgical Education

In an era of rapid technological advancement and evolving healthcare landscapes, surgical education is faced with a myriad of challenges that must be addressed to ensure the competence and confidence of future surgeons while guaranteeing the safety of patients[24,25]. Today, surgical residents and educators must overcome challenges related to high-stress environments in and out of the operating room, patient safety concerns, varying exposure and experience, and limited feedback[21,26,27,24].

Given the high-stakes environment and technical skills involved in surgery, surgical training fosters a high-stress environment[26]. Unlike more common and less technically demanding procedures, learning complex surgical operations becomes more challenging and stressful due to the increased risk of patient harm.[26] Not only does this put the surgical educator in a difficult position, balancing the responsibilities of teaching the surgical trainee and maintaining patient safety, but it also makes acquiring the technical skills necessary for such procedures more difficult for the trainee. [21,26]

Additionally, varying exposure poses an issue among surgical trainees[28,29]. Exposure relies on the surgical cases available, which can be unpredictable in terms of duration and frequency[28,29]. In more specialized areas of training in both neurosurgical and orthopedic spine surgery, case availability greatly varies depending on the residency program, resulting in limited opportunities for some surgical residents to acquire the appropriate technical skills[28,29]. This limitation has led trainees to work on days off in order to meet training requirements, leading to increased stress and feelings of burnout; these phenomena indicate that inadequate training may contribute to concerns about career development and burnout[26]. At the same time, the introduction of reduced hours to address burnout issues has further decreased learning opportunities for surgical trainees[30]. Moreover, varying exposure presents a complex issue in surgical education.

Finally, gaining feedback is another challenge in surgical education[27]. Positive and negative feedback is a critical component in surgical training and education, as it allows the learner to understand the composites of expertise and how to acquire technical skill sets during their training[31]. While it is a requirement for surgical educators to provide feedback to their trainees, meaningful, postoperative feedback tends to be given irregularly[27]. However, this is largely due to surgical instructors' demanding schedules and responsibilities[21,27]. Solutions are needed to accommodate the learning needs of surgical trainees and the demanding schedule and responsibilities of surgical educators.

Such challenges in surgical education are well-documented, and measures are being taken through research, educators, and policymakers to ensure the proper education of surgical residents. Surgical education is shifting towards a competency-based quantitative training model to address these complex issues[7,8].

Shift Towards Competency-Based Training

To address key challenges in surgical education and the evolving field of medicine, medical and surgical education have shifted towards implementing a competency-based model into training[7,8]. This model's learning objectives are centered on competence, or how well learners can accomplish a task, rather than time[7]. As a result, this framework ensures the safety of patients and uniform educational objectives and competence across training programs. Competency-based assessments have infiltrated surgical training in several ways.

A defining step towards a competency-based framework in surgical education is the introduction of Entrusted Professional Activities (EPAs) into surgical training[32]. EPAs are tasks or responsibilities that can be entrusted to an unsupervised trainee after showing sufficient competence over several occasions[32,33]. Such tasks and responsibilities range from technical to interprofessional skills, covering all roles of the surgical profession. EPAs create structure within the traditional apprenticeship model and enable key learning components, such as discussion, assessment, and feedback to be easily incorporated into the curriculum[32]. Moreover, by standardizing competency-based learning objectives in surgical training, EPAs ensure patient safety and quality outcomes[32]. However, while EPAs construct a buildable framework for competency-based education, challenges related to its effective execution remain, including uniform implementation across residency programs and the lack of science that guides the direction and implementation of EPAs[34]. In Canada, post-graduate training is based on a physician competency framework called CanMEDS[35]. In this framework, a physician is considered a medical expert with 6 intrinsic roles, communicator, collaborator, leader, advocate, scholar, and professional[35]. Further, in Canadian post-graduate training, EPAs are composed of 5 to 15 milestones that are associated with the CanMEDS roles[35].

The shift towards competency-based training has required a redefining of the focus of trainee assessment among surgical educators and researchers[36]. A common and widely accepted performance assessment tool in surgical education research, evaluation, and training is the objective structured assessment of technical skills (OSATS)[37]. This subjective Likert scale assessment comprises several items reflecting technical surgical skills that surgical educators use to evaluate their trainees, including hemostasis, respect for tissue, instrument handling, economy of movement, flow, knowledge of procedure, and an overall rating[37]. The hemostasis item refers to the ability to control bleeding while respect for tissue relates to the ability to avoid and minimize potential harm to surrounding anatomical structures[38]. Instrument handling relates to the surgeons' or trainees' ability to effectively use instruments, and the economy of movement refers to the extent to which repetitive, non-purposeful movements are made[38]. Flow refers to the forward planning of an operation, reflected by a seamless transition between steps and movements of a procedure[38]. Knowledge of procedure assesses the trainee's understanding of the entire procedure including the steps, instruments, and relevant anatomy[38]. OSATS is considered the gold standard in surgical evaluations and is often modified to meet the assessment goals of the specialty and operation[37]. While OSATS serves as a standardized assessment tool, studies have recently criticized its inability to oversee all aspects of surgical training[39,40].

Lastly, the current surgical curriculum requires surgical residents to enter their cases into the Accreditation Council for Graduate Medical Education (ACGME) resident case log system. Graduating surgical residents must enter 750 major operative cases with at least 150 entered during their final year[5,41]. However, as requested by the ACGME, residents may only log an operation when the individual has played a significant role in five competencies: diagnosis, preoperative care, operation selection, operation, and postoperative care[42]. To be actively and consistently involved in such aspects of an individual patient is highly unlikely given the limited autonomy and condensed surgical rotations residents experience[42]. Such measures were put in place to ensure surgical residents experience appropriate breadth and depth of surgical operations[41].

The movement toward competency-based training facilitates standardized competence across surgical programs and their respective trainees[7]. Not only does this contribute to increased patient safety, but it can also contribute to reduced burnout among surgical residents[7,43,44]. The latter can be explained by competency-based training's ability to offer equal training opportunities, translating to an increased level of readiness[44]. Through standardizing assessment and creating milestones, surgical residents can feel more confident in their operative abilities, a major component of reducing burnout[43,44].

While competency-based training strives for standardized opportunity and competence among surgical trainees, case availability remains an issue in surgical education[45]. Exposure to specific surgical cases varies per residency program[39,40,46]. For example, spine operations are specialized neurosurgical and orthopedic residency procedures[39,40,46]. Moreover, depending on the hospital and program, exposure and experience to spine surgery greatly varies among surgical residents[29,39,40,46]. While curricular measures can be put in place to ensure all residents gain exposure to a specific operation, case availability is the limiting factor to this, which can make such a curricular objective difficult to achieve[45]. Surgical simulation may be a valuable tool in tackling this challenge and contribute to the shift toward competency-based training in surgical education.

Surgical Education through Simulation Training

Modern surgical education is progressively incorporating simulation training, a method of learning by practicing clinical skills in a simulated environment[9]. Simulation can come in various modes including virtual reality, which is primarily used for technical skills, and simulated standardized patients, which is utilized for practicing skills like diagnostics[9]. Moreover, simulation training could be useful for formative (focused on progress and learning) and summative (focused on certifying competency) assessments in surgical education, which are integral components of competency-based training[9].

Simulation training in surgical education has many advantages. In simulated environments, surgical trainees can obtain knowledge and focus on specific skills, whether technical, interprofessional, or behavioral[9]. Most notably, simulation provides a controlled, risk-free environment where surgical trainees can devote themselves to learning essential clinical skills without putting patient safety and quality of care at risk[14]. Further, surgical trainees can master skills and make errors without the stress and potential harm associated with learning in the operating room[9]. This translates to a better understanding of when errors can take place and instigates the development of mitigation and prevention strategies for such errors; gaining this skill in a simulated environment enhances the surgical trainees' readiness as independent surgeons while ensuring patient safety[14]. Simulation training also allows trainees to be exposed to clinical variation, a typically difficult aspect to control for and include during clinical rotations[9]. This aspect contributes to greater breadth and standardized competence across surgical programs.

Scientific data supporting simulation training in surgical education continues to emerge, highlighting its ability to develop diverse aspects of clinical skills for trainees[47]. While simulation training provides a valuable platform for surgical trainees to master and acquire clinical skills, thorough research is essential to ensure its effective implementation and that trainees fully benefit from the simulation experience[47]. Specifically, validation studies are essential for understanding the educational utility of simulators in surgical training.

Validation

As simulation gains traction in surgical education, validation emerges as a pivotal foundational phase in assessing the effectiveness of simulators for surgical training[10]. Validation studies aim to understand the appropriateness of a tool for a particular goal[11]. For example, in the context of surgical simulation, a validation study for a surgical simulator would aim to understand its utility as a learning tool in surgical training. Currently, surgical simulation literature primarily follows a traditional framework while educational theorists accept a contemporary framework[11]. Both frameworks are outlined in Table 1.

Table 1: Validation Frameworks

Framework	Approach	
Traditional	Establish concepts of validity.	**Face**: the extent to which the simulator replicates the real procedure
		Content: the extent to which the simulator measures the skills they were designed to simulate
		Construct: the ability of the simulator to distinguish different operative skill levels and includes convergent validity
		Predictive: the extent to which the simulator can predict future performance, especially that of the operating room
Messick's Contemporary	Construct a validity argument by gathering evidence of validity from up to five sources.	**Test Content**: the relationship between a tool's content and the construct it aims to measure
		Response Process: the integrity of the data collection
		Internal Structure: the measures taken to determine the degree to which items of an instrument align with the underlying construct and are reported as statistical measures
		Relationship to Other Variables: the degree of relatedness between assessment measures and external independent measures
		Consequences: the potential and observed consequences of the tool of interest

The traditional framework of validation involves "types of validity", including face, content, construct, and transfer or concurrent validity[11]. These types of validity can be divided into two approaches: subjective and objective validation[10,48]. Subjective validation utilizes expert opinion to determine the value of the examined instrument, and it involves face and content validity[10,11]. In the context of surgical simulation, face validity refers to the extent to which the simulator visually resembles the surgical task, while content validity refers to the extent to which the simulator's surgical task reflects that of the surgical task done in real life[10,11,48]. These types of validity require expert input, and while subjective questionnaires are typically administered for assessing face and content validity, a universal consensus of evaluation does not exist[10,11,48,49]. Objective validation involves using experimental means to ascertain the extent to which the simulator's surgical task parallels the same task performed in the operating room[10,48]. Notable objective validation measures include construct validity and transfer or concurrent validity. Construct validity evaluates the simulator's ability to differentiate skill levels in the surgical task[10,11]. To assess this, experimental studies examine the performance of trainees compared to that of expert surgeons on the simulator of interest. Convergent validity is a subset of construct validity that examines how closely measures of the same construct agree with another[50,51]. This is often evaluated by investigating the extent of agreement between a targeted measure and a well-known measure[50,51]. Further, this validity is suggestive of the simulator's utility by relating its performance assessment with that of what is used in surgical training. Transfer or concurrent validity refers to the extent to which the simulator can predict future performance, especially that of the operating room[10,48]. This type of validity typically involves longitudinal studies to understand the transfer of skill from simulation to an accepted "testing" task like using ex vivo tissues or cadavers[10,48]. Moreover, predictive validity is typically assessed after determining the face, content, and construct validity of a simulator. This framework of validity is used extensively in surgical simulation literature, although a contemporary framework is accepted in the education community[11,52].

Messick's contemporary framework of validity proposes that validity is an argument consisting of an accumulation of evidence that supports a tool's use for a particular purpose and population[37]. It postulates that all evidence of validity relates to construct and comes from five sources, content, response process, internal structure, relation to other variables[37], and consequences. The "test content" dimension refers to the relationship between a tool's content and the construct it aims to measure[11,52]. This source of evidence must be based on the input from participants who are experts in the procedure of interest[11,52]. "Response process" pertains to the integrity of the data collection, including standardized instructions and blinded raters[11,52]. "Internal structure" relates to the measures taken to determine the degree to which items of an instrument align with the underlying construct and are reported as statistical measures such as internal consistency and reliability[11,52]. The "relationship to other variables" dimension refers to the degree of relatedness between assessment measures and external independent measures such as proficiency level and experience[11,52]. Finally, the "consequences" concept refers to evidence relating to the potential and observed consequences of the tool of interest.

Messick's contemporary framework is the recommended approach in educational research, as advocated by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), in *Standards for Educational and Psychological Testing*[11,52]. However, the integration of this approach into surgical education research has been slow[11,52,53]. A study from 2018 found that only 6.6% of validation studies for surgical simulation from 2008 to 2017 used the contemporary framework[52]. This trend is speculated to occur to maintain consistency among past literature[11].

The traditional and contemporary frameworks are formally distinct. Noticeably, the contemporary framework focuses on gathering evidence compared to establishing validation as in the traditional framework. The contemporary framework also values implementing research methods to enhance the quality of validation studies, evident in the "response process" and "internal structure" criteria. Nonetheless, a significant overlap exists between the two approaches[11]. Specifically, "face validity" and "content validity" in the traditional framework are tightly related to the contemporary framework's "test content". In addition, the traditional framework's "construct validity", including "convergent validity", is virtually the same as the contemporary framework's "relationship to other variables" aspect. This trend follows the traditional framework's "predictive validity" which relates to "consequences" in the contemporary framework. Moreover, because establishing validity principles plays a critical step in evaluating the utility of simulation in surgical training, a compromise between the frameworks involving clear definitions and justifications of validity methods may be the most practical way forward in future simulation validation studies[11].

This study investigates the foundational steps involved in validation studies, namely establishing face, content, construct, and convergent validity. Establishing such principles sets the groundwork for future studies that outline a tool's role in surgical training.

Virtual Reality Spine Simulation

Surgical simulation is becoming an important tool in surgical training for technical skills, with laparoscopic surgery being one of the most advanced areas[54]. In the United States, surgical simulation is implemented in laparoscopic training and assessment of performance[54]. Virtual

Reality simulation is an emerging tool in surgical education, although its application to spine surgery is minimal[55]. In past years, studies have evaluated the utility of virtual reality spine simulators with many focusing on the pedicle screw insertion technique [15,55]. Nonetheless, only a limited number of spine surgery pedicle screw insertion simulation platforms exist; however, they lack comprehensive validity and high fidelity, highlighting the need for the development of more pertinent simulation training tools[15,16,17,18,19].

However, despite the growing number of virtual reality spine simulation studies, recent reviews from Pfandler et al. and McCloskey et al. have determined that the majority of these platforms are limited in quality based on scoring using the Medical Education Research Study Quality Instrument and the GRADE criteria, respectively[15,55]. Further, although current literature points to promising uses of virtual reality surgical simulation, the lack of robust literature on virtual reality spine simulation has limited its adoption in spine surgery training[20-22]. Consequently, such reviews advocate for future studies to assess how training on virtual reality spine simulators demonstrates skill transfer to the operating room[15,55]. Other notable suggestions for future virtual reality spine simulator studies include justified, validated, and reliable metrics, and clinical expert ratings in their assessment[15,55]. Considering these aspects in future virtual reality spine simulation studies would increase the credibility of implementing virtual reality simulation in spine surgery training. Virtual reality simulation for spine surgery training may be an important advancement in surgical education, as it addresses the challenges that residents face regarding restrictions and limitations in clinical hours[22]. Moreover, for virtual reality simulation to be implemented into spine surgery training, comprehensive studies must be carried out with relevance to the operating room.

Pedicle Screw Insertion and Its Associated Risks

The pedicle screw insertion is a common, widely used technique in spine surgery. This is utilized in procedures like scoliosis, spine tumors, trauma, infection, and degenerative disease[56]. The procedure involves creating an entry point on the vertebral body using an awl followed by preparing a channel using a cannulation probe, otherwise known as a pedicle finder, that advances through the vertebral cancellous bone[57]. At this point, the surgeon largely depends on tactile feedback and experience-based judgment to determine the location of the channel.[56] To identify any errors in channel preparation, a ball tip probe is inserted into the channel, where the surgeon feels for any breaches that may have been made in the process[57]. The channel is pre-threaded using a tap before further breach verification with a ball tip probe and insertion of the screw. Final X-Rays can be performed to ensure the proper positioning of the screw.[56]

While performing these steps, the surgeon must utilize the limited spinal anatomical landmarks, which are subject to morphological variability, to make informed decisions on the accuracy and safety of the procedure[57]. This aspect becomes crucial given this technique's limited margin of error, as the pedicle is close to many vital neural and vascular structures[56,57]. Today, image-guided techniques are employed in place to prevent the malplacement of screws, including fluoroscopy, intraoperative navigation, and robotic assistance[56]. Despite the advancements in navigational aid, mastering the pedicle screw insertion technique remains crucial, as resources at hospitals vary and technical disruptions can make navigational aids unavailable.

Pedicle screw insertions pose risks for complication if not inserted correctly. For example, although rare, malpositioned screws can put surrounding neural and vascular structures at serious risk of damage, including complications like dysesthesia, hemorrhage, and neurological injury.[57,58] Suboptimal positioned screws can also lead to early construct failure or pseudoarthrosis formation.[57] Moreover, the potential harm associated with the malplacement of pedicle screws is well documented with an incidence ranging between 4.2-7.8%.[58,59] The pedicle screw insertion proves to be complex and demanding, necessitating a steep learning curve.[56] Recent publications showed that trainees need to place 60 to 80 pedicle screws under direct supervision before being able to independently perform accurate and safe pedicle screw insertions[12,13]. With varying exposure, limited cases, and restricted hours,[22,29,60] such a degree of experience may be difficult to achieve for training neurosurgical and orthopedic residents and spine fellows. Furthermore, tools for comprehensive surgical training could be valuable in gaining the technical skills necessary for mastering the pedicle screw insertion technique.

TSYM Simulator

The TSYM simulator is a non-immersive virtual reality platform developed by Cedarome Canada Inc. dba Symgery. (Montreal, Canada). This system provides various simulated surgical scenarios, primarily focusing on spine interventions. The TSYM simulator is a stand-alone system, consisting of a screen that displays the 3D surgical environment, a robotic arm attached to the operative tool, and three tool handles for simulating an array of surgical instruments, as seen in Figure 1. The simulator utilizes a voxel-based system to achieve a realistic intra-operative user experience, enabling haptic feedback during the simulated operations. A previous study examining the utility of virtual reality simulation in surgical training suggests that such simulators with haptic feedback result in increased accuracy in cervical pedicle screw insertions compared to training through traditional means[61]. Moreover, the simulator's tactile feedback coupled with audio feedback enhances the fidelity of the simulator's surgical tasks. The TSYM simulator creates a non-immersive operative environment, whereby the simulated procedure is limited to the screen, unlike immersive virtual reality platforms that provide a 360-degree virtual environment. Although a non-immersive platform may possess lower fidelity compared to an immersive platform, a recent study comparing the effectiveness of immersive and non-immersive virtual reality training for hip arthroscopy found similar outcomes related to skill and procedural acquisition and skill transfer[62]. Further, such features of the TSYM simulator make it a more promising tool for surgical training.

The TSYM simulator offers an L4-L5 bilateral pedicle screw insertion scenario, an essential technique in spine surgery with a steep learning curve[12,13]. The following manuscript aims to establish the foundational principles of the L4-L5 bilateral pedicle screw insertion scenario, investigating face, content, construct, and convergent validity. To our knowledge, this is the first study to assess convergent validity for an L4-L5 bilateral pedicle screw insertion on a virtual reality platform.

STUDY RATIONALE, HYPOTHESIS, AND OBJECTIVES

Rationale

Surgical training involves acquiring complex, bimanual skills while ensuring patient safety under a stressful and high-stakes environment. Such challenges become heightened in spine surgery training where mastering technical skills is critical, exposure in residency varies, and the need for comprehensive training is essential[29,60,8]. Virtual reality simulators may be a valuable tool to overcome such issues, as they provide residents with practical and accessible training in a safe, stress-free environment.

However, simulation has not been implemented into training for spine surgery, as current simulators lack comprehensive validation studies, preventing the uptake into surgical training. To address the challenges in teaching spine surgery among neurosurgical and orthopedic residents, we aimed to validate the utility of a virtual reality spine simulator's lumbar pedicle screw insertion scenario, a critical skill in spine surgery with a steep learning curve. In this study, a consensus approach between the traditional and contemporary validation frameworks was used to evaluate the simulator's educational potential, where components of the traditional framework were evaluated to construct a validity argument.

Hypothesis

The TSYM virtual reality simulator's L4-L5 bilateral pedicle screw insertion scenario will demonstrate face, content, construct, and convergent validity, contributing to evidence of validity of the simulator's potential as a formative tool in spine surgery training.

Objectives

The objectives of this case series study are:

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 of the simulation's performance metrics by assessing the relationship between the simulation-derived metrics and simulated pedicle screw insertion operative performance OSATS.
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.

MANUSCRIPT

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

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

[5] Department of Neurosurgery, Hamilton General Hospital, McMaster University Medical Centre, Hamilton, Ontario, Canada

[6] Department of Neurosurgery, Zayed Military Hospital, Abu Dhabi, United Arab Emirates

The preceding work has been augmented with additional information and materials to reflect the requirements for thesis submission for a Master of Science.

Manuscript submitted for review to the Journal of Neurosurgery (June 18th, 2024).

INTRODUCTION

Surgical training involves balancing the objectives of imparting complex skills and ensuring patient safety[25]. Intraoperative surgical teaching offers personalized instruction but may involve limited exposure to complex procedures with the potential for patient harm[63,64]. This becomes particularly relevant in spine surgery, where mastery of technical skills is essential, exposure in residency varies, and the need for comprehensive training is essential[29,60,8]. Pedicle screw insertion is a common but technically demanding spine surgical procedure[8,57]. Mastering the pedicle screw insertion involves a steep learning curve since trainees need to place many pedicle screws under direct supervision before being able to independently perform safe pedicle screw placement[12,13]. The potential harm associated with pedicle screw insertion malposition is well documented, and in two large literature review articles, the incidence of pedicle screw malposition ranges between 4.2 – 7.8%[58,59].

The role of virtual reality simulation in enhancing surgical education and providing a risk-free environment for procedural learning and skill refinement continues to develop[57,65,66]. There are a limited number of spine surgery pedicle screw insertion simulation platforms. Many lack comprehensive validity and high fidelity, highlighting the need for the development of more relevant simulation training tools[15,16,17,18,19]. The need to shift towards quantitative competency-based surgical education is becoming increasingly clear[8]. This would standardize training methods, focusing on the development and assessment of specific competencies rather than using time in training as an indicator of experience[67]. Such standardization is important in complex surgical procedures like pedicle screw insertions, where competency of specific skills directly impacts patient outcomes[58,59].

The TSYM Symgery virtual reality platform allows for a realistic pedicle screw insertion simulation and provides personalized feedback. This system provides an array of performance metrics useful to assess surgical techniques, offering an innovative approach to surgical training[68-70]. The educational utility of the TSYM Simulator platform is yet to be established. This study explores the simulator's training potential by gathering subjective and objective validity evidence, specifically face, content, construct, and convergent validity[10,50,71]. Face validity refers to the extent to which the simulator replicates the real procedure while content validity refers to the extent to which the simulator measures the skills they were designed to simulate[10,48]. Face and content validity can be determined through questionnaires[48]. Construct validity is a type of objective validity that describes the ability of the simulator to distinguish different operative skill levels and can be investigated by comparing surgical performance between "less skilled" and "skilled" groups[71,48,72]. Simulation-derived performance metrics on tool handling and the Objective Structured Assessment of Technical Skills (OSATS) ratings, the gold standard for scoring performance in surgical education in human operative procedures, were used to assess construct validity[73,74]. Convergent validity, a subgroup of construct validity, explores the degree of agreement between different measures of the same construct and is typically evaluated by correlating the measure of interest to a well-known measure [50,51]. We examine convergent validity by investigating how well the simulation-derived performance metrics relate to OSATS[50,51].

Gallagher and co-workers have reviewed and outlined fundamental principles of the traditional framework of validation by applying scientific methods for the assessment of surgical education and training[10]. Messick's contemporary framework of validity proposes that validity is an argument consisting of an accumulation of evidence that supports a tool's use for a particular purpose and population[37]. This study aims to utilize both methods to gather evidence of validity for the utilization of the TSYM simulator platform in spine surgical training. This approach may potentially provide a more holistic evaluation of the TSYM systems' capacity to assess and train learners in complex procedures like the pedicle screw insertion simulation[10,48]. 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, non-spine 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. Participants were categorized a priori into two groups, skilled participants (Post Graduate 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. After signing the consent, participants completed a demographic questionnaire. Participants were then provided with standardized written and verbal instructions regarding the steps and instruments available to complete the simulated L4-L5 bilateral pedicle screw insertion on the TSYM simulator. Verbal and written instructions were administered in English; however, given the bilingualism presence in Quebec, language-related questions, specifically any French-related questions or issues, were welcomed and answered appropriately by an on-site individual involved in running the trial. Participants then performed a dry lab and an L2 simulated laminectomy procedure to become acquainted with the TSYM simulator and simulated tools and their functions (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, and once completed, required participant confirmation before proceeding. This article follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines[75].

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). The three-dimensional (3D) intraoperative spinal surgical procedures present in this simulator rely on a voxel-based system[72] (Figure 1B). The simulator consists of a single haptic arm that provides continuous tactile feedback during operator manipulation of the surgical instruments employed to complete the task (Figure 1C) and generates appropriate auditory and visual information for each tool used. This system is equipped with a robust software platform including pre-programmed surgical tools and captures multiple performance metrics, enabling a detailed analysis of surgical performance. The pedicle screw insertion simulation task consists of 1 animated and 4 deconstructed interactive steps described in Table 2. These steps were repeated for each screw. For standardization purposes, users performed the pedicle screw insertions using constant magnification and inserted 6.5 x 45 mm pedicle screws in a predetermined order left L5, left L4, right L5, right L4, (see supplemental information). Participants had access to live X-rays to verify the entry point and angles for pedicle cannulation and confirm inserted screw accuracy. Video 1 shows a skilled participant performing a pedicle screw insertion on the simulator.

Face and Content Validity

The neurosurgical and orthopedic spine surgeons and spine 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 [72,76]. A consensus on an acceptable median value for sufficient face and content validity has not been established[72,76]. Since no gold standard exists for face and content validity, in this study, the overall simulated procedure and its deconstructed tasks were considered to have adequate evidence of face and content validity if questionnaires achieved a median ≥ 4.0 on the 7-point Likert scale, consistent with our previous studies[72,76].

Construct Validity

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

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 on 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 2. The rationale to treat each pedicle screw insertion by each participant independently was that each screw insertion involved a different simulated vertebrae entry point, orientation, and angulation.

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[29,30]. 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 5 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 (α = .97 [95% CI, .96, .98]) and excellent inter-rater reliability (α = .97 [95% CI, .97, .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 was imported into Python to develop tool metrics. Outliers in tool metrics were identified and imputed on MATLAB R2023b. All other statistical assessments were performed on

SPSS (version 29.0; IBM, Armonk, New York). The data was not normally distributed as assessed by Shapiro-Wilk's test ($P < .05$). Mann-Whitney tests assessed statistical differences between groups for each performance measure. A two-tailed Spearman Rank Order Correlation Coefficient examined associations between performance metrics.

RESULTS

Participants

Demographic data and relevant information concerning the two groups in this case series study are presented in Table 3. A total of 27 participants from two Quebec universities were included in this investigation. 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 independently. The difference between the two groups was statistically significant, ($P < .001$). 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 4. The 4 participating spine surgeons and 2 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. Related to content validity, all steps achieved adequate evidence of validity (median $\geq$ 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 5). Significant differences were found between the two groups in 4 of 25 performance metrics. According to how convergent validity is assessed in studies in the literature, there is a documented anticipated result[50,51]. We therefore anticipated 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[77-79]. 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 (.0014, 95% CI [.00119, .00153] vs .001, 95% CI [.0012, .0013]; P =.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]; P = .01). Although the 3D acceleration values were small across both groups, statistical analysis confirmed a significant difference (P = .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]; P = .04) and spent significantly more time in contact with surrounding tissue than the skilled group (.22, 95% CI [.18, .25] vs .11, 95% CI [.09, .13]; P <.001). These group 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 ($P < .001$ for each item). 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 = -.32, $P < .01$; Spearman's Coefficient = -.39, $P < .01$; Spearman's Coefficient = -.37, $P < .01$; Spearman's Coefficient = -.38, $P < .01$; Spearman's Coefficient = -.29, $P < .01$; Spearman's Coefficient = -.33, $P < .01$, 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 = -.25, $P < .01$; Spearman's Coefficient = -.34, $P < .01$; Spearman's Coefficient = -.42, $P < .01$; Spearman's Coefficient = -.43, $P < .01$; Spearman's Coefficient = -.31, $P < 0.01$; Spearman's Coefficient = -.31, $P < .01$, respectively). No significant correlations were found between the tap's 3D velocity and 3D acceleration and OSATS items. Table 6 outlines the associations between these performance metrics.

DISCUSSION

The results of this case series study may be useful for surgical educators and researchers interested in spine simulation for several reasons. First, the pedicle screw insertion simulation employed in this investigation demonstrated varying degrees of validity: mixed and variable levels of face and content, as well as mixed evidence of construct and convergent validity. These subjective and objective results contribute to the evidence of validity as an argument for this platform's potential as a formative educational tool in spine surgery training[25]. Second, to our knowledge, this is the first study to correlate simulator-derived metrics with OSATS ratings to assess the convergent validity of a simulated operative procedure on a virtual reality spine surgery platform. Third, using OSATS ratings in simulator performance assessment and simulator-derived metrics provides a more holistic understanding of learner operative performance. This methodology may be useful to investigators interested in designing and validating simulators focused on improving technical skills during surgical training.

Face, Content, and Construct Validity

The traditional validation framework investigates types of validity like face, content, and construct; while, the contemporary framework gathers evidence from up to five sources (content, response process, internal structure, relation to other variables, and consequences) to support a tool's use for a particular purpose and population[37]. This study combines both frameworks, using traditional types of validity to help construct a validity argument for the TSYM simulator's educational utility in surgical training. This validity argument is primarily supported by the OSATS findings and rather weakly by the other validity measures. Moreover, as elaborated below, the validity argument lacks strength and would benefit from more robust findings.

The participating spine surgeons and fellows rated most face and content validity statements with a median of 4.0 or greater, which is considered to provide adequate evidence of face and content validity[72,76]. While these results are consistent with our definition of "adequate" face and content validity, this evidence can be considered "mixed" for two reasons. First, we did not anticipate participants providing a rating of "totally realistic" (7) and our group has, accordingly, previously considered a median of "4" as sufficient for providing evidence of face and content validity[72,76]. Second, the broad ranges of observed ratings of most items, some including "1" and "7", illustrate meaningful variance within the experienced participants' perspectives. Participants were asked to comment on the simulator's L4-L5 pedicle screw insertion scenario. Verbal feedback from this group indicated that torque feedback utilizing the tap for pre-threading the inner pedicle canal could be improved to enhance the realism of this step with the lowest median value. These results are suggestive of borderline reasonable face validity and content validity; however, because of the great variability, the results must be interpreted with care. The L4-L5 pedicle screw simulation will need to be improved to enhance its realism.

The study demonstrated statistically significant differences between the two groups for four simulation-derived tool metrics of 25 using two tools: 3D velocity and 3D acceleration of the simulated tap, and the maximum force and the tool contact of the simulated screwdriver (Figure 2). The skilled group had higher 3D velocity than the less skilled associated with tap screw use. The skilled group's familiarity with the procedural components[77] and operative technical skills needed may allow this group to use increased velocity using the simulated tap. The less skilled group being less experienced and more hesitant in the use of this instrument may have resulted in lower tap velocity. The skilled group, conscious of the safety risk of high acceleration instrument usage, may utilize lower tap acceleration consistent with previous studies highlighting that experience in pedicle screw fixation is an important factor distinguishing participant expertise[12,13]. The maximum force applied by the screwdriver was significantly higher for the less skilled group than the skilled group consistent with previous virtual reality studies assessing instrument force application[77-79]. Studies using artificial neural networks (ANN) were able to assess junior and senior residents, neurosurgeons, and orthopedic surgeons' performance and identify different patterns of force application, which is considered a safety metric[79-81]. From a clinical standpoint, increasing the force applied can result in breaches in the medial, lateral, and upper and lower vertebral directions. This could place many neurological and vascular structures, such as the adjacent nerve root, the dura, and arteries and veins at the anterior component of the vertebral column, at risk of injury. Our results involving maximum force applied by the screwdriver are consistent with a pattern of force application in which more skilled groups appreciate that using high forces during screwdriver use may impact patient safety and therefore moderate this metric during their training and career[79]. A different pattern may be the reason why the less skilled group had higher screwdriver tool contact. The less skilled group may be more unsure concerning appropriate screwdriver application and use on the pedicle screw due to lesser anatomical and practical knowledge of the procedure, resulting in more inadvertent adjacent tissue contact.

Only four of 25 tool-related performance metrics provided evidence of construct validity. The limited number of significant metrics identified could be related to the low number of participants in the study. The possibility exists that less skilled individuals trained to modify these metrics to more closely correspond to those of skilled participants may improve their operative performance. However, the identification of these four metrics allowed further studies to assess the convergent validity of the simulation platform.

The skilled group significantly outperformed the less skilled group in each OSATS component (Figure 3). These OSATS studies support the evidence of simulator-derived instrument tool metrics validation concerning the construct validity of the TSYM simulator for the L4-L5 pedicle screw insertion simulation.

Correlating Simulation-Derived Performance Metrics and OSATS Ratings for Convergent Validity

The ability to correlate novel simulation-derived metrics with OSATS scoring allowed an assessment of the convergent validity of the TSYM platform[29,30]. The finding that two of four simulation-derived performance metrics correlated with all OSATS items provided evidence of convergent validity for the TSYM simulator· and has several implications. The OSATS ratings of participant video pedicle screw insertion performance identified that screwdriver maximum force application and screwdriver tool contact were negatively correlated with all OSATS items. The less skilled groups' OSATS ratings for pedicle screw insertion were significantly lower, consistent with their results on these two simulation-derived metrics discussed previously. Two of the four significant simulation-driven performance metrics, 3D velocity, and 3D acceleration using the tap, did not significantly correlate with the OSATS ratings. This finding may suggest that these performance features are not accurately captured in the items rated by OSATS. This may relate to the expert evaluators scoring these videos' inability to visually accurately determine these specific two composites of expertise, 3D velocity and 3D acceleration of tap instrument while in the bone channel[79,82]. Although OSATS is a validated method to assess surgical performance, several studies have questioned the ability of OSATS to fully measure the complexities of surgical operating room performance[39,40]. This study suggests that the consideration of utilizing OSATS and other surgeon educator assessments of surgical performance in combination with those provided by simulator-derived metrics may enhance our understanding, assessment, and training of surgical skills and be useful for formative assessment. Integration of these two methodologies may result in a more comprehensive assessment of learner surgical expertise.

These studies allow further investigations related to the predictive validity of the TSYM simulator. This would necessitate that participants' results, obtained from their simulated performance on the TSYM simulator with pedicle screw insertion, would predict their future pedicle screw insertion performance on human patients.

TSYM as an Educational Tool

The result of this investigation suggests that certain aspects of TSYM simulator pedicle screw insertion scenario may be useful for training less skilled learners. Specifically, trainees having access to performance ratings on the 4 metrics, which provided evidence of construct validity, may improve their pedicle screw insertion results. Virtual reality simulators have been assessed in pedicle screw placement training and have improved the accuracy of screw placement[8,60,70,83]. A study investigating simulation training has shown its utility in accelerating skill acquisition in pedicle screw placement[37]. Less skilled trainees may benefit from incorporating virtual reality simulation for performing complex spine procedures into the spine surgery learning curriculum and as a potential formative educational tool[69,70]. While specific features of the TSYM simulator pedicle screw insertion scenario may be useful, this simulation platform may need modification to meet its full potential as a surgical educational system.

With the vast data generated from virtual reality simulators like the TSYM platform, artificial intelligence methodologies may be useful for enhancing the understanding of the precision and granularity of surgical skills[84]. Artificial neural networks can utilize this data to identify new metrics and rank their importance in simulated operative performance helping surgical educators focus on critical metrics for gaining specific operative technical skills[79,80,83]. The availability of simulated pedicle screw operative performance data from novices and experts and the utilization of deep learning algorithms can be used to create intelligent tutoring systems like the Intelligent Continuous Expertise Monitoring System (ICEMS) developed by our group[42,84]. However, artificial intelligence-enhanced curriculum can be associated with unintended outcomes, and care is required in developing programs necessitating human educator input[85]. Deep learning applications utilizing simulator-derived metric results and the equivalent OSATS video ratings for each procedure may allow future artificial intelligence systems to predict OSATS scoring utilizing only the evaluation of the simulator-derived metrics.

One objective of virtual reality studies is to combine artificial intelligence approaches, which can assess human instrument tracking data critical to optimal operative performance[33]. This data along with OSATS ratings and intelligent tutoring systems can be incorporated into a human "Intelligent Operating Room" that could possess the ability to continually assess and train learners while minimizing surgical errors [76,82,83,86].

Limitations

The TSYM simulation platform has limitations. First, the pedicle screw insertion simulation does not capture the dynamic intraoperative environment consisting of the learner and surgical educator providing continuous personalized feedback. Second, the simulated procedure was developed with one animated and 4 deconstructed steps in a linear, unidirectional sequence of pedicle screw insertions, which does not represent the flexible approach available during human pedicle screw insertion procedures. Third, the TSYM simulator consists of a single-handed robotic arm setup, which does not reproduce the bimanual psychomotor skills utilized during patient spinal procedures[40,69,82]. This study included neurosurgeons and orthopedic surgeons focused on spine surgery, as well as fellows, and neurosurgical and orthopedic residents. While significant attempts were made to increase the participant pool, the scheduling of participants due to respective clinical commitments limited the number of study participants, thereby limiting the generalization of results. The small sample size also meant that statistical analyses for construct and convergent validity were underpowered, meaning that some significant differences may be the result of a type 1 error. While a common limitation in surgical education studies, especially with medical residents, fellows, and surgeons, future studies must include larger numbers of skilled and less skilled participants from multiple institutions to improve the robustness of results and generalizability[84]. In this study, each pedicle screw insertion was evaluated individually due to differences in entry points, screw angulation, and anatomy. Larger studies will be necessary to evaluate the impact of repeated pedicle screw insertion on the learning curves of skilled and less skilled groups associated with this simulated procedure. To standardize the pedicle screw insertion procedure a fixed-size screw was utilized, however, the TSYM platform offers a wide variety of screw sizes and lengths to assess learners' ability to perform these procedures. While PGY5-6 residents and non-spine fellows possess significantly greater anatomical and practical knowledge in pedicle screw insertions, these study participants outlined high variability in prior experience with this technique. This variability could contribute to the limitations in the findings, particularly in distinguishing performance differences in the other metrics assessed. Future studies should determine skill groupings based on experience, such as including a pre-requisite number of screws for each group. Finally, because the study was administered in English, language barriers could have affected the clarity of instructions for some participants, which could have limited the participant's performance on the simulated task. Future Canadian studies should provide an option for all instructions to be administered in both French and English.

CONCLUSION

While several limitations and challenges exist with the TSYM simulator platform pedicle screw insertion scenario, some aspects of this simulator's scenario, such as performance metrics of screwdriver maximum force and screwdriver tool contact, show potential to assist in surgical teaching. Information garnered from this study may allow improvements in the TSYM simulator so that it can be even more useful in this regard in the future.

THESIS DISCUSSION

Contributions to Original Knowledge

This study contributes to the surgical education literature, specifically concerning gaining evidence of validity for surgical virtual reality spine simulators, in the following ways:

1. To our knowledge, this study is the first time in which OSATS have been employed for determining construct validity of a virtual reality spine simulator platform for simulated pedicle screw insertion, and
2. To our knowledge, this investigation is the first to utilize the evaluation of convergent validity to provide evidence for the validity of a virtual reality spine simulator.

Validity Evidence

The validation study combines Messick's contemporary framework and the traditional framework of validity. While the traditional validity types are evaluated in this study, the implications of the findings are viewed as an attempt to find evidence for constructing a validity argument, supporting the educational utility of the TSYM virtual reality simulator in surgical training.

This validation study can be viewed through the lens of Messick's contemporary validity framework. As previously mentioned, Mesick's contemporary validity framework involves accumulating evidence of validity from five sources: test content, response process, internal structure, relations to other variables, and consequences. In this study, neurosurgical and orthopedic spine surgeons and spine fellows rated statements related to the content of the pedicle screw insertion simulation using assessed with a 7-point Likert scale[72,76]. All but one statement was deemed adequate; however, the results should be viewed with caution given the variability of responses. This measure meets the "content" criteria of Messick's contemporary validity framework, whereby the content of the simulated task aligns with the components and skills of the real procedure. Additionally, the study included measures to reduce bias in the assessment process including standardized verbal and written instructions, uniform steps and tools, and randomized-blinded rating. These efforts to maintain the integrity of the data constitute gaining "response process" evidence. The validation study also gathered "internal structure" evidence, which relates to the measures taken to explore the reliability of scores to measure the same construct, often through statistical means. Specifically, this study evaluated the OSATS ratings' inter-rater reliability and internal consistency, which resulted in excellent values. Finally, the validation study demonstrated a "relationship to other variables" by observing significant group differences in OSATS ratings and simulation-derived metrics. The significant correlation between two simulation-derived metrics and all OSATS ratings also contributes to this avenue of evidence. However, the study was not designed to gather evidence of validity relating to Messick's "consequences" concept, which entails the potential and actual consequences related to the assessment tool. Moreover, this study was able to gather evidence from four out of five sources of validity, supporting the TSYM simulator's educational potential in surgical training.

Future Directions

Surgical Simulation Timeline

The implementation of simulators into surgical residency training follows a methodological timeline. Surgical simulators must undergo several steps of validation, involving thoroughly planned research studies[10]. The initial phases of validation include establishing features involving visual and methodological realism of the simulated procedure and the capability of discriminating skill proficiency[48]. Following this phase, investigations directly related to surgical trainees' learning can be performed[48]. Such studies increase the understanding of a simulator's potential role in surgical training[48].

This study demonstrates mixed and variable evidence for face, content, construct, and convergent validity of the TSYM simulator's L4-L5 bilateral pedicle screw insertion. These results provide some evidence of the educational potential of TSYM simulator's L4-L5 bilateral pedicle screw insertion for surgical training. All the data outlined in this study will be provided to the manufacturer to help the engineers involved improve the educational utility of the simulator. The study serves as an important assessment of the utility of the L4-L5 pedicle screw insertion scenario on the TSYM simulator, paving the way for future modifications and improvements of the simulator. More investigations will be essential to further evaluate its educational utility, including skill development, training methods, and clinical implications.

Future studies related to the TSYM simulator's L4-L5 bilateral pedicle screw insertion should be carried out to greater understand its implications in surgical training. These studies should be longitudinal and track the progress of surgical trainees to reflect and investigate the simulator's role as a formative training tool. Future studies should provide targeted feedback, as this is a crucial component in learning and skill development. The incorporation of such features enables the generation of learning curves that can increase the understanding of its impact as a training tool. Finally, determining skill transfer from the simulator to real operations is instrumental in elucidating the simulator's role in surgical training. Such a study would more clearly identify the simulator's utility and its clinical implications.

As mentioned previously, the study has other implications, related to the simulator's ability to produce large amounts of data. The TSYM simulator generated 3D reconstructions of inserted pedicle screws within the vertebra. This data can be used to evaluate more clinically relevant aspects of surgical performance such as entry points, screw angles, and breaches. Because this study was able to establish a degree of construct validity, surgical performance data can be assessed with artificial intelligence algorithms to uncover the granularity of surgical skills, such as identifying critical features of performance[80,84]. Such findings can contribute to enhancing surgical education, as surgical educators can focus on teaching these skill features to trainees. Artificial intelligent tutors can also be developed, which provide continuous personalized feedback during the simulated procedure and tailored feedback after the simulated procedure completion. These systems may identify weaknesses in learner technical skills and provide feedback on how to avoid errors and improve performance[83]. However, future studies should assess the impact of teaching skill features identified by artificial intelligence to understand the varying effects such methodology can have[85]. These research avenues can contribute toward the shift to competency-based training to the development of quantitative assessment and training curriculum development.

THESIS SUMMARY

Surgical education is shifting from an apprenticeship framework to competency-based quantifiable frameworks. While this transition addresses several challenges in surgical training, it requires tools that can accurately and continuously quantify the expertise composites of surgical performance. Virtual reality simulators provide a safe and risk-free environment for developing critical and technically challenging realistic scenarios which can assess and train learners to acquire the psychomotor technical skills required for mastery of operative performance. This case series investigation demonstrates that the pedicle screw insertion simulation employed demonstrated varying degrees of validity: mixed levels of face and content, as well as mixed evidence of construct and convergent validity. This evidence may help contribute to the validity argument for this platform's potential as a formative educational tool in spine surgery training. However, the variability in the median response of the spine fellows and spine surgeons in terms of face and content validity, the fact that only 4 of 25 performance metrics significantly discriminated skilled from less skilled surgeons, and the mixed evidence of construct and convergent validity, suggest that the true value of the TSYM simulator's L4-L5 pedicle screw insertion at its current form must be interpreted with caution. Improvements in the simulator and/or scenario will be needed to allow it to meet its full potential as a surgical teaching tool. To our knowledge, this is the first investigation to assess the convergent validity of a simulated operative procedure on a virtual reality spine surgery platform by correlating simulator-derived metrics and OSATS ratings. The utilization of OSATS ratings in simulator performance assessment together with simulator-derived metrics may be useful to researchers interested in designing and validating simulators and curricula focused on improving technical skills during surgical training.

Pedicle screw insertions are a common yet technically challenging skill for stabilizing the spine in neurosurgery and orthopedic surgery[8,57]. However, mastery of this technique involves a steep learning curve with trainees needing to practice between 60 to 80 screws with direct supervision to be able to independently perform pedicle screw insertions accurately and safely[12,13]. Technical errors in this procedure may cause significant patient harm, posing high risks when acquiring the skillsets for this technique[58]. Virtual reality surgical simulators may be a valuable, risk-free tool in developing technical operative skills, like pedicle screw insertion[57,65,66].

This case series study investigated the potential educational utility of a simulated L4-L5 bilateral pedicle screw insertion on the TSYM virtual reality spine simulator study to gather validity evidence. The objectives of the study were to 1) evaluate face and content validity for an L4-L5 bilateral pedicle screw insertion simulation on the TSYM simulator platform, 2) use simulation-derived metrics and the assessment of simulated pedicle screw insertion operative performance utilizing OSATS to assess construct validity, 3) establish convergent validity of the simulation's performance metrics by assessing the relationship between the 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.

The TSYM simulator's L4-L5 bilateral pedicle screw insertion demonstrated emerging face, content, construct, and convergent validity. The simulated procedure's visual and content-related realism was considered adequate based on the inputs of participating spine fellows and surgeons. However, due to the variability in median responses (ranging from 1.0 to 7.0), the true adequacy of face and content validity must be interpreted with caution. Related to construct validity, significant group differences were only found in 4 out of 25 simulation-derived performance metrics assessed. However, significant group differences were consistent among OSATS ratings, as the skilled group significantly outperformed the less skilled group in each OSATS item and the overall OSATS rating. Finally, 2 out of 4 simulation-derived performance metrics significantly negatively correlated with each OSATS item and the overall rating. The two significant negative correlations were consistent with convergent validity, as the finding matched the predicted relationship. The varying degree of consistency related to construct validity and the limited number of participants cautions against the generalizations of the study's findings, hence, the results are considered mixed.

The validity evidence gathered in this study lays the groundwork for understanding the educational utility of the TSYM simulator's L4-6 bilateral pedicle screw insertion and the aspects needing improvement. The findings of this study may help to begin to construct a validity argument supporting the TYSM's potential as a formative training tool for surgical training. However, the strength of this argument should be interpreted with caution given the various limitations highlighted throughout the thesis. Future studies are required to elucidate its learning potential, impact on surgical proficiency, and clinical implications.

In summary, this case series study suggests that the TYSM simulator's L4-L5 bilateral pedicle screw insertion scenario has some degree of educational potential for skill development among surgical trainees, but improvements are needed to optimize this potential. Virtual reality simulators capable of replicating pedicle screw insertions, like the TSYM simulator (but improved based upon research studies like the one presented here), may be useful in surgical education, as they provide a safe, risk-free environment for surgical trainees to focus and develop essential and technically challenging operative skills.

REFERENCES

1. Allen L, O'Connell A, Kiermer V. How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. *Learned Publishing*. 2019;32(1):71-74. doi:https://doi.org/10.1002/leap.1210

2. Brand A, Allen L, Altman M, Hlava M, Scott J. Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing*. 2015;28(2):151-155. doi:https://doi.org/10.1087/20150211

3. Anderson O, Davis R, Hanna GB, Vincent CA. Surgical adverse events: a systematic review. *Am J Surg*. Aug 2013;206(2):253-62. doi:10.1016/j.amjsurg.2012.11.009

4. Stulberg JJ, Huang R, Kreutzer L, et al. Association between surgeon technical skills and patient outcomes. *JAMA surgery*. 2020;155(10):960-968.

5. Camison L, Brooker JE, Naran S, Potts JRI, Losee JE. The History of Surgical Education in the United States: Past, Present, and Future. *Annals of Surgery Open*. 2022;3(1):e148. doi:10.1097/as9.0000000000000148

6. Polavarapu HV, Kulaylat AN, Sun S, Hamed OH. 100 years of surgical education: the past, present, and future. *Bull Am Coll Surg*. Jul 2013;98(7):22-7.

7. Skjold-Ødegaard B, Søreide K. Competency-based Surgical Training and Entrusted Professional Activities – Perfect Match or a Procrustean Bed? *Annals of Surgery*. 2021;273(5):e173-e175. doi:10.1097/sla.0000000000004521

8. McGaghie WC. Mastery learning: it is time for medical education to join the 21st century. *Acad Med*. Nov 2015;90(11):1438-41. doi:10.1097/ACM.0000000000000911

9. Bjerrum F, Thomsen ASS, Nayahangan LJ, Konge L. Surgical simulation: Current practices and future perspectives for technical skills training. *Medical Teacher*. 2018/07/03 2018;40(7):668-675. doi:10.1080/0142159X.2018.1472754

10. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surgical Endoscopy And Other Interventional Techniques*. 2003/10/01 2003;17(10):1525-1529. doi:10.1007/s00464-003-0035-4

11. Hughes T, Fennelly JT, Patel R, Baxter J. The Validation of Surgical Simulators: A Technical Report on Current Validation Terminology as a Reference for Future Research. *Cureus*. Nov 2022;14(11):e31881. doi:10.7759/cureus.31881

12. 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*. Jun 2012;21(6):1151-6. doi:10.1007/s00586-011-2065-2

13. Gonzalvo A, Fitt G, Liew S, et al. The Learning Curve of Pedicle Screw Placement: How Many Screws Are Enough? *Spine*. 2009;34(21):E761-E765. doi:10.1097/BRS.0b013e3181b2f928

14. Marshall SD, Nataraja RM. Patient Safety and Surgical Education. In: Nestel D, Dalrymple K, Paige JT, Aggarwal R, eds. *Advancing Surgical Education: Theory, Evidence and Practice*. Springer Singapore; 2019:327-337.

15. Pfandler M, Lazarovici M, Stefan P, Wucherer P, Weigl M. Virtual reality-based simulators for spine surgery: a systematic review. *Spine J*. Sep 2017;17(9):1352-1363. doi:10.1016/j.spinee.2017.05.016

16. Luciano C, Banerjee P, Florea L, Dawe G. Design of the immersivetouch: a high-performance haptic augmented virtual reality system. 2005:

17. Shi J, Hou Y, Lin Y, Chen H, Yuan W. Role of Visuohaptic Surgical Training Simulator in Resident Education of Orthopedic Surgery. *World Neurosurg*. Mar 2018;111:e98-e104. doi:10.1016/j.wneu.2017.12.015

18. Xin B, Chen G, Wang Y, et al. The Efficacy of Immersive Virtual Reality Surgical Simulator Training for Pedicle Screw Placement: A Randomized Double-Blind Controlled Trial. *World Neurosurg*. Dec 28 2018;doi:10.1016/j.wneu.2018.12.090

19. Chen T, Zhang Y, Ding C, et al. Virtual reality as a learning tool in spinal anatomy and surgical techniques. *N Am Spine Soc J*. Jun 2021;6:100063. doi:10.1016/j.xnsj.2021.100063

20. 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)*. Dec 1 2018;15(6):677-685. doi:10.1093/ons/opy030

21. Gasco J, Patel A, Ortega-Barnett J, et al. Virtual reality spine surgery simulation: an empirical study of its usefulness. *Neurol Res*. Nov 2014;36(11):968-73. doi:10.1179/1743132814Y.0000000388

22. Wang Z, Shen J. Simulation Training in Spine Surgery. *J Am Acad Orthop Surg*. May 1 2022;30(9):400-408. doi:10.5435/jaaos-d-21-00756

23. Peracchia A. Presidential Address: Surgical education in the third millennium. *Ann Surg*. Dec 2001;234(6):709-12. doi:10.1097/00000658-200112000-00001

24. Kotsis SV, Chung KC. Application of the "See One, Do One, Teach One" Concept in Surgical Training. *Plastic and Reconstructive Surgery*. 2013;131(5):1194-1201. doi:10.1097/PRS.0b013e318287a0b3

25. Leung A, Luu S, Regehr G, Murnaghan ML, Gallinger S, Moulton C-a. "First, Do No Harm": Balancing Competing Priorities in Surgical Practice. *Academic Medicine*. 2012;87(10):1368-1374. doi:10.1097/ACM.0b013e3182677587

26. Robinson DBT, Luton O, Mellor K, et al. Trainee perspective of the causes of stress and burnout in surgical training: a qualitative study from Wales. *BMJ Open*. 2021;11(8):e045150. doi:10.1136/bmjopen-2020-045150

27. Nathwani JN, Glarner CE, Law KE, et al. Integrating Postoperative Feedback Into Workflow: Perceived Practices and Barriers. *Journal of Surgical Education*. 2017/05/01/ 2017;74(3):406-414. doi:https://doi.org/10.1016/j.jsurg.2016.11.001

28. Daniels AH, Alsoof D, McDonald CL, et al. Longitudinal Assessment of Modern Spine Surgery Training: 10-Year Follow-up of a Nationwide Survey of Residency and Spine Fellowship Program Directors. *JB JS Open Access*. Jul-Sep 2023;8(3)doi:10.2106/jbjs.Oa.23.00050

29. Daniels AH, Ames CP, Garfin SR, et al. Spine surgery training: is it time to consider categorical spine surgery residency? *The Spine Journal*. 2015/07/01/ 2015;15(7):1513-1518. doi:https://doi.org/10.1016/j.spinee.2014.08.452

30. Antiel RM, Reed DA, Van Arendonk KJ, et al. Effects of Duty Hour Restrictions on Core Competencies, Education, Quality of Life, and Burnout Among General Surgery Interns. *JAMA Surgery*. 2013;148(5):448-455. doi:10.1001/jamasurg.2013.1368

31. Fadelalla MG, Elbakri S, Poon M. Intraoperative feedback: A survey of surgical trainees' perspective. *Scott Med J*. May 2023;68(2):58-62. doi:10.1177/00369330231163375

32. Tobin S. Entrustable Professional Activities in Surgical Education. In: Nestel D, Dalrymple K, Paige JT, Aggarwal R, eds. *Advancing Surgical Education: Theory, Evidence and Practice*. Springer Singapore; 2019:229-238.

33. Ten Cate O. Nuts and bolts of entrustable professional activities. *J Grad Med Educ*. Mar 2013;5(1):157-8. doi:10.4300/jgme-d-12-00380.1

34. Kitto S, Fantaye AW, Zevin B, Fowler A, Sachdeva AK, Raiche I. A Scoping Review of the Literature on Entrustable Professional Activities in Surgery Residency Programs. *Journal of Surgical Education*. 2024/06/01/ 2024;81(6):823-840. doi:https://doi.org/10.1016/j.jsurg.2024.02.011

35. LoGiudice AB, Sibbald M, Monteiro S, et al. Intrinsic or Invisible? An Audit of CanMEDS Roles in Entrustable Professional Activities. *Academic Medicine*. 2022;97(8):1213-1218. doi:10.1097/acm.0000000000004731

36. Szasz P, Grantcharov T. The role of assessment in surgical education. *Advancing Surgical Education: Theory, Evidence and Practice*. 2019:221-228.

37. Martin JA. Measuring the Impact of Educational Interventions: A Quantitative Approach. In: Nestel D, Dalrymple K, Paige JT, Aggarwal R, eds. *Advancing Surgical Education: Theory, Evidence and Practice*. Springer Singapore; 2019:389-403.

38. Niitsu H, Hirabayashi N, Yoshimitsu M, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surg Today*. Mar 2013;43(3):271-5. doi:10.1007/s00595-012-0313-7

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. *Journal of Surgical Education*. 2016/11/01/ 2016;73(6):1020-1025. doi:https://doi.org/10.1016/j.jsurg.2016.04.023

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. *Clinical Orthopaedics and Related Research®*. 2016/04/01 2016;474(4):874-881. doi:10.1007/s11999-015-4603-4

41. Sachdeva AK, Bell RHJ, Britt LD, Tarpley JL, Blair PG, Tarpley MJ. National Efforts to Reform Residency Education in Surgery. *Academic Medicine*. 2007;82(12):1200-1210. doi:10.1097/ACM.0b013e318159e052

42. Morgan R, Kauffman DF, Doherty G, Sachs T. Resident and Attending Perceptions of Resident Involvement: An Analysis of ACGME Reporting Guidelines. *Journal of Surgical Education*. 2017/05/01/ 2017;74(3):415-422. doi:https://doi.org/10.1016/j.jsurg.2016.10.012

43. Engelhardt KE, Bilimoria KY, Johnson JK, et al. A National Mixed-Methods Evaluation of Preparedness for General Surgery Residency and the Association With Resident Burnout. *JAMA Surgery*. 2020;155(9):851-859. doi:10.1001/jamasurg.2020.2420

44. Shaw RD, Lamb CR, Carroll MC, Wong SL, Rosenkranz KM. Implementation of a Standardized Sub-Intern Curriculum Improves Confidence for those Entering a Surgical Residency. *Journal of Surgical Education*. 2022/11/01/ 2022;79(6):1402-1412. doi:https://doi.org/10.1016/j.jsurg.2022.07.002

45. Vinden C, Malthaner R, McGee J, et al. Teaching surgery takes time: the impact of surgical education on time in the operating room. *Can J Surg*. Apr 2016;59(2):87-92. doi:10.1503/cjs.017515

46. Daniels AH, Ames CP, Smith JS, Hart RA. Variability in Spine Surgery Procedures Performed During Orthopaedic and Neurological Surgery Residency Training: An Analysis of ACGME Case Log Data. *JBJS*. 2014;96(23):e196. doi:10.2106/jbjs.M.01562

47. de Montbrun SL, Macrae H. Simulation in surgical education. *Clin Colon Rectal Surg*. Sep 2012;25(3):156-65. doi:10.1055/s-0032-1322553

48. 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*. Mar 2010;24(3):536-46. doi:10.1007/s00464-009-0634-9

49. Huang C, Cheng H, Bureau Y, Agrawal SK, Ladak HM. Face and content validity of a virtual-reality simulator for myringotomy with tube placement. *J Otolaryngol Head Neck Surg*. Oct 20 2015;44:40. doi:10.1186/s40463-015-0094-2

50. Fried MP, Sadoughi B, Weghorst SJ, et al. Construct Validity of the Endoscopic Sinus Surgery Simulator: II. Assessment of Discriminant Validity and Expert Benchmarking. *Archives of Otolaryngology–Head & Neck Surgery*. 2007;133(4):350-357. doi:10.1001/archotol.133.4.350

51. 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. *Journal of Surgical Education*. 2022/07/01/ 2022;79(4):1000-1008. doi:https://doi.org/10.1016/j.jsurg.2022.01.014

52. Borgersen NJ, Naur TMH, Sørensen SMD, et al. Gathering Validity Evidence for Surgical Simulation: A Systematic Review. *Annals of Surgery*. 2018;267(6):1063-1068. doi:10.1097/sla.0000000000002652

53. American Educational Research Association APA, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing. *Standards for Educational and Psychological Testing*. 2014 ed. American Educational Research Association; 2014.

54. Aggarwal R. Simulation in Surgical Education. In: Nestel D, Dalrymple K, Paige JT, Aggarwal R, eds. *Advancing Surgical Education: Theory, Evidence and Practice*. Springer Singapore; 2019:269-278.

55. McCloskey K, Turlip R, Ahmad HS, Ghenbot YG, Chauhan D, Yoon JW. Virtual and Augmented Reality in Spine Surgery: A Systematic Review. *World Neurosurg*. May 2023;173:96-107. doi:10.1016/j.wneu.2023.02.068

56. Perna F, Borghi R, Pilla F, Stefanini N, Mazzotti A, Chehrassan M. Pedicle screw insertion techniques: an update and review of the literature. *MUSCULOSKELETAL SURGERY*. 2016/12/01 2016;100(3):165-169. doi:10.1007/s12306-016-0438-8

57. Manbachi A, Cobbold RS, Ginsberg HJ. Guided pedicle screw insertion: techniques and training. *Spine J*. Jan 2014;14(1):165-79. doi:10.1016/j.spinee.2013.03.029

58. 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*. Oct 2011;31(4):E8. doi:10.3171/2011.7.FOCUS11168

59. Hicks JM, Singla A, Shen FH, Arlet V. Complications of pedicle screw fixation in scoliosis surgery: a systematic review. *Spine (Phila Pa 1976)*. May 15 2010;35(11):E465-70. doi:10.1097/BRS.0b013e3181d1021a

60. Grantcharov TP, Reznick RK. Teaching procedural skills. *BMJ*. May 17 2008;336(7653):1129-31. doi:10.1136/bmj.39517.686956.47

61. Hou Y, Shi J, Lin Y, Chen H, Yuan W. Virtual surgery simulation versus traditional approaches in training of residents in cervical pedicle screw placement. *Arch Orthop Trauma Surg*. Jun 2018;138(6):777-782. doi:10.1007/s00402-018-2906-0

62. Rahman OF, Kunze KN, Yao K, et al. Hip Arthroscopy Simulator Training With Immersive Virtual Reality Has Similar Effectiveness to Nonimmersive Virtual Reality. *Arthroscopy*. Mar 19 2024;doi:10.1016/j.arthro.2024.02.042

63. Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. *Acad Med*. Aug 2003;78(8):783-8. doi:10.1097/00001888-200308000-00006

64. Rattner DW, Apelgren KN, Eubanks WS. The need for training opportunities in advanced laparoscopic surgery. *Surg Endosc*. Oct 2001;15(10):1066-70. doi:10.1007/s004640080021

65. 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*. Dec 2015;22(6):636-42. doi:10.1177/1553350615579729

66. Rogers MP, DeSantis AJ, Janjua H, Barry TM, Kuo PC. The future surgical training paradigm: Virtual reality and machine learning in surgical education. *Surgery*. 2021/05/01/ 2021;169(5):1250-1252. doi:https://doi.org/10.1016/j.surg.2020.09.040

67. Bissonnette V, Mirchi N, Ledwos N, Alsidieri G, Winkler-Schwartz A, Del Maestro RF. Artificial Intelligence Distinguishes Surgical Training Levels in a Virtual Reality Spinal Task. *J Bone Joint Surg Am*. Dec 4 2019;101(23):e127. doi:10.2106/jbjs.18.01197

68. Sewell C, Morris D, Blevins NH, et al. Providing metrics and performance feedback in a surgical simulator. *Comput Aided Surg*. Mar 2008;13(2):63-81. doi:10.3109/10929080801957712

69. 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):1-7.

70. 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*. May 2015;10(5):603-18. doi:10.1007/s11548-014-1091-z

71. Huang C, Cheng H, Bureau Y, Ladak HM, Agrawal SK. Automated Metrics in a Virtual-Reality Myringotomy Simulator: Development and Construct Validity. *Otology & Neurotology*. 2018;39(7):e601-e608. doi:10.1097/mao.0000000000001867

72. 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)*. Dec 15 2020;20(1):74-82. doi:10.1093/ons/opaa269

73. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med*. Dec 1996;71(12):1363-5. doi:10.1097/00001888-199612000-00023

74. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg*. Jul 2010;97(7):972-87. doi:10.1002/bjs.7115

75. 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*. Apr 2008;61(4):344-9. doi:10.1016/j.jclinepi.2007.11.008

76. 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)*. Jan 8 2024;doi:10.1227/ons.0000000000001044

77. 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*. Aug 2021;35(8):4417. doi:10.1007/s00464-020-08038-w

78. Sawaya R, Bugdadi A, Azarnoush H, et al. Virtual Reality Tumor Resection: The Force Pyramid Approach. *Oper Neurosurg (Hagerstown)*. Jun 1 2018;14(6):686-696. doi:10.1093/ons/opx189

79. Mirchi N, Bissonnette V, Ledwos N, et al. Artificial Neural Networks to Assess Virtual Reality Anterior Cervical Discectomy Performance. Article. *Operative Neurosurgery*. Jul 2020;19(1):65-75. doi:10.1093/ons/opz359

80. Alkadri S, Ledwos N, Mirchi N, et al. Utilizing a multilayer perceptron artificial neural network to assess a virtual reality surgical procedure. *Computers in Biology and Medicine*. 2021/09/01/ 2021;136:104770. doi:https://doi.org/10.1016/j.compbiomed.2021.104770

81. Reich A, Mirchi N, Yilmaz R, et al. Artificial Neural Network Approach to Competency-Based Training Using a Virtual Reality Neurosurgical Simulation. Research Support, Non-U.S. Gov't. *Operative Neurosurgery*. 07 01 2022;23(1):31-39. doi:https://dx.doi.org/10.1227/ons.0000000000000173

82. 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(2):e0229596. doi:10.1371/journal.pone.0229596

83. 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*. 2022/04/26 2022;5(1):54. doi:10.1038/s41746-022-00596-8

84. 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(8):e198363-e198363. doi:10.1001/jamanetworkopen.2019.8363

85. 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(9):e2334658-e2334658. doi:10.1001/jamanetworkopen.2023.34658

86. 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(2):e2149008-e2149008. doi:10.1001/jamanetworkopen.2021.49008

APPENDIX

Supplemental Digital Content 1. Methods. Simulated L4 & L5 pedicle screw placement scenario

The TSYM Symgery platform is a virtual reality simulator platform with one haptic arm and a number of interchangeable handles, including a Kerrison and a straight handle. Participants performed two tasks before proceeding with the pedicle screw insertion 1) a Dry Lab which was followed by 2) a L2 laminectomy simulation scenario to become acquainted with the TSYM simulator the simulated instruments and their function.

The Dry Lab involved an interactive display of instrument handling utilizing the haptic handle. Participants used the straight handle to perform the following tasks: 1) creating a hole utilizing the awl, 2) removing a spherical object with the burr, and 3) creating a trajectory using the pedicle finder. Participants then were asked to utilize the simulated Kerrison handle to bit off three simulated bony areas.

When the Dry Lab is completed successfully participants are given verbal instructions on the performance of the L2 laminectomy procedure that they will be asked to complete and provided with written information concerning each step of the procedure. The L2 simulation includes 1 animated and 4 interactive steps. The animated step begins with a pre-exposed surgical cavity with the spinous process and the interspinous ligaments removed from the simulated patient's spine. The first interactive step involved the use of the 4mm burr to thin the L2 lamina by removing the cancellous bone component. In the second interactive step the ligamentum flavum was detached using an angled curette, in the third interactive step a 4mm Kerrison was used to remove the remaining lamina and resect the detached ligament flavum. Once the participant is satisfied with the decompression, the fourth interactive step follows which involves utilizing a Woodson to verify the complete bilateral removal of the ligamentum flavum.

After completing the Dry Lab and L2 laminectomy participants are then provided with verbal and written instructions on how to perform the L4 & L5 pedicle screw insertion placement simulation.

This simulation also starts with an animated component outlining the L4 & L5 vertebrae being completely dissected from a posterior approach. The standardized screen magnification was maintained for all participants and a specific order for screw placement was outlined. This involved beginning with the left L5 screw, followed by the left L4, then the right L5 and concluding with the right L4. Each step was associated with a restricted list of simulated instruments which participants had to pick before moving to the next step. Participants started at left L5, creating an entry point with the awl. Live fluoroscopy was available during the procedure to verify the entry point, insertion angulation and screw placement. The next step was to create a channel in the pedicle utilizing the pedicle finder. Then, a 2 mm ball tip probe was used to check for any evidence of a pedicle breach. The participant must declare the presence of a breach from an automatic prompt before moving to the next step. The screw channel was then tapped using a 5.5 mm tap, and the 2 mm ball tip probe was used once again to check for any possible breach. The last step involved inserting a standardized to 6.5 mm x 45 mm simulated pedicle screw. On completion of each screw insertion, the simulator created a 3D model, illustrating the individuals' placed screw placement. The final information available to the participant involved a 3D reconstruction of each of the 4 pedicle screws along with written feedback on the participant's overall performance.

Figures

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 provide 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 available to mimic an array of tools in the virtual environment.

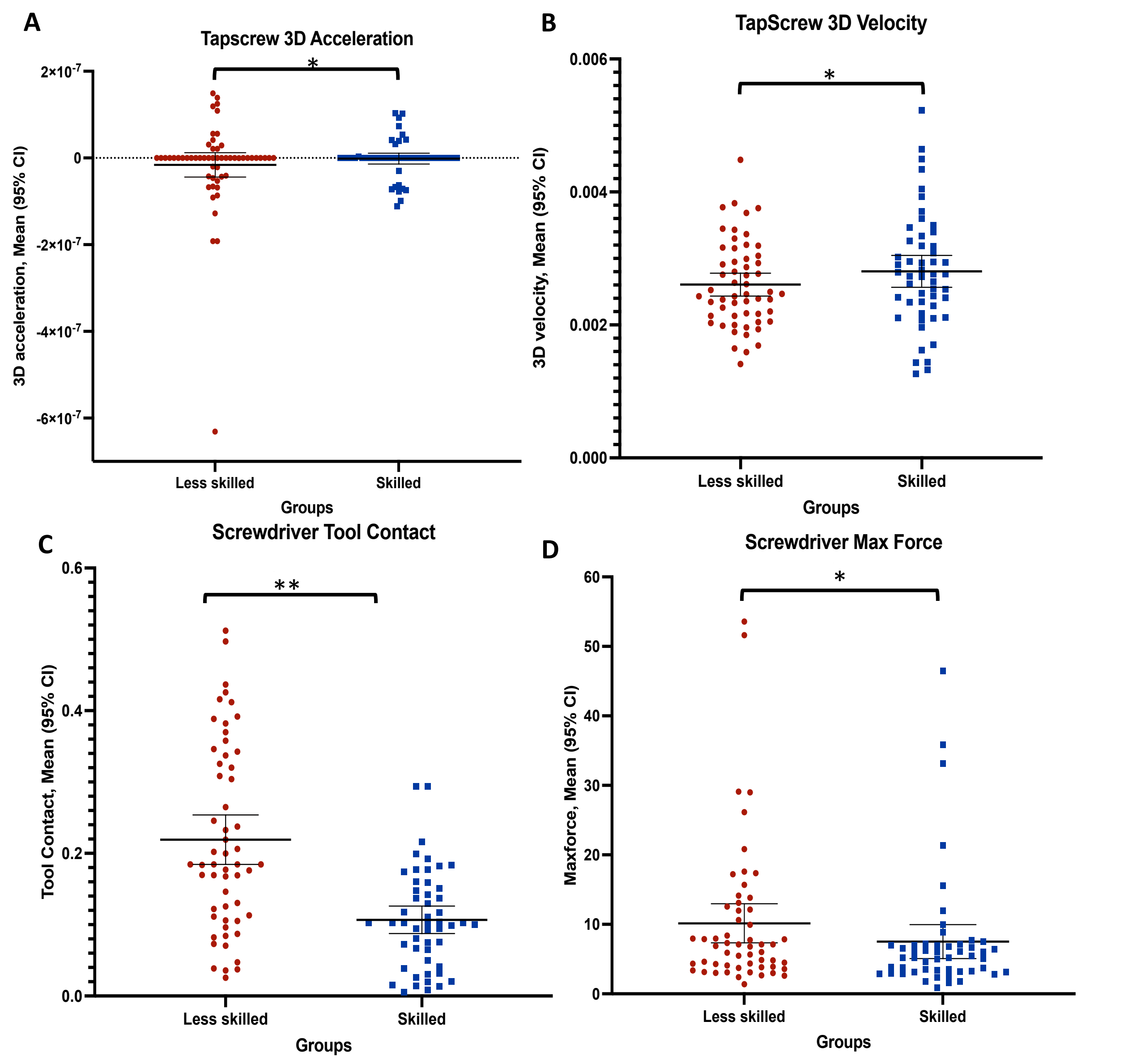

Figure 2. Significant Performance Assessments of the Task Using Simulation-Generated Performance Metrics. **A**, Tap screw's 3D Velocity. **B**, Tap screw's 3D Acceleration. **C**, Screwdriver Max Force on the pedicle. **D**, Screwdriver Contact with pedicle. The central line indicates the mean value for each group. *Represents a significant difference between groups after Mann-Whitney U, nonparametric test ($p < .05$). **Represents a significant difference between groups after Mann-Whitney U, nonparametric test ($p < .01$).

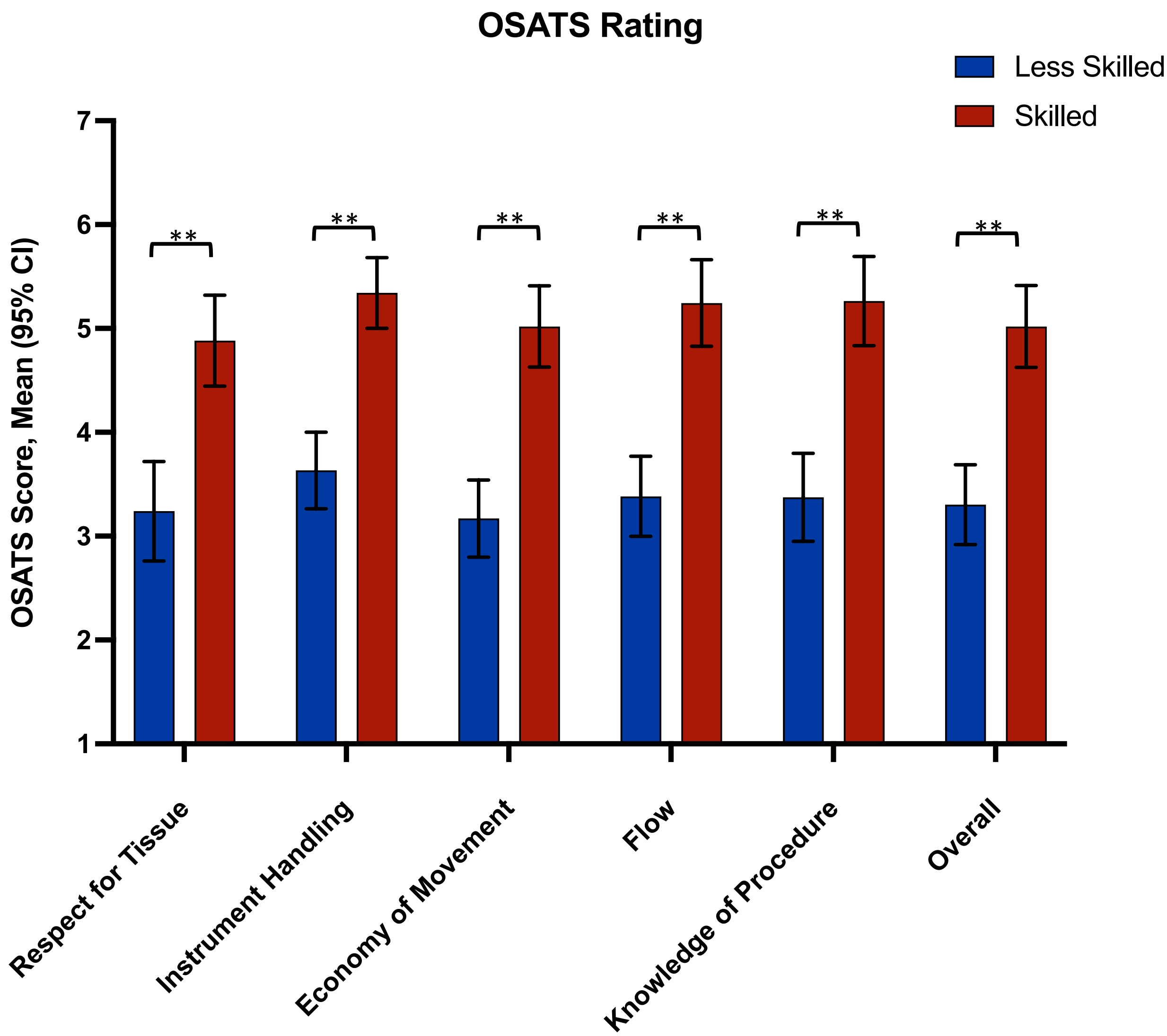

Figure 3. Performance Assessment of the Pedicle Screw Insertion Task Using OSATS. *Represents a significant difference between groups after Mann-Whitney U, nonparametric test ($p < .05$). **Represents a significant difference between groups after Mann-Whitney U, nonparametric test ($p < .01$).

Tables

Table 2: 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 2: Channel Creation	Create channel in the pedicle and verification using fluoroscopy	Pedicle finder
Step 3: Channel Breach Verification	Check for presence or absence of a pedicle breach	2mm ball tip probe
Step 4: Tap Insertion	Pre-thread the previously created channel in the pedicle and verification using fluoroscopy	5.5mm tap
Step 5: Pedicle Breach Verification	Check for presence or absence of a pedicle breach	2mm ball tip probe
Step 6: Screw insertion	Insertion of the selected screw by rotation the screwdriver and verify using fluoroscopy	Screwdriver and Screw (6.5 mm diameter and 4.5mm length)

Table 3: 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 5-6	-	-
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 = Post Graduate Year

SD = Standard Deviation

**No significant difference was found between the two groups except for the mean number of reported pedicle screws inserted. ($P < .001$)

Table 4: 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 colour 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.

Table 5: Simulation-Derived Metrics Obtained from the L4-L5 bilateral Pedicle Screw Insertion Simulation on the TSYM Simulator and Corresponding Mann-Whitney U 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, nonparametric test ($P < .05$).

Table 6: Concurrent 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	-0.01	0.90	0.06	0.54	0.11	0.28	0.09	0.34	0.01	0.88	0.01	0.89
Tap 3D Acceleration	-0.17	0.09	-0.12	0.21	-0.18	0.07	-0.15	0.12	-0.13	0.19	-0.14	0.16

*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.