THE DEVELOPMENT AND VALIDATION OF THE LUMBAR MICRODISCECTOMY ASSESSMENT SCALE

By

Dr. Ganeshwaran Shivapathasundram

A thesis submitted in conformity with the requirements for the degree of

Master of Philosophy

Sydney Medical School

University of Sydney

Table of Contents

STATEMENT OF ORIGINALITY

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Dr Ganeshwaran Shivapathasundram

ACKNOWLEDGEMENT

I am deeply grateful to my primary supervisor, Dr. Pramudith Sirimanna, whose unwavering support was essential to the completion of this thesis. Pram's guidance and inspiration were instrumental throughout this project. I am particularly thankful for his consistent leadership, patience, perseverance, and even temperament over the past several years.

I also extend my sincere thanks to Dr. Sarah Aitken and Professor Vasi Naganathan for their excellent supervision, guidance, and mentorship during my Master of Philosophy candidature. Their willingness to offer their time and expertise, often on short notice, was invaluable, and their wisdom is deeply appreciated.

Above all, I must express my profound gratitude to my wife, Laura, and my three children. Laura's unwavering support has been the foundation upon which this work was built. Balancing a clinical practice, research, and a young family would have been impossible without her enduring support. For my children, hopefully this thesis contributes to the advancement of medical education, which I hope you one day benefit from. Finally, I dedicate this thesis to my parents. I am eternally grateful to the opportunities they have provided me through their own hard work, dedication and sacrifice. Without them none of this would be possible.

ABSTRACT

Lumbar microdiscectomy (LM) is the index spinal surgical operation for neurosurgery and orthopaedic surgery trainees. Surgical training has moved towards competency-based training, yet there is no objective competency-based assessment tool for LM. In this thesis, we reviewed the literature on educational tools in spinal surgery. While we identified a number of models with which to educate spinal surgery trainees, at this stage, there are no procedure-specific assessment tools in spinal surgery. In order to develop an objective evaluation tool for assessment of LM, the procedure was broken down into its composite steps using a hierarchical task analysis. We then performed a Delphi analysis with both international and local neurosurgeons working in teaching hospitals to reach expert consensus on the lumbar microdiscectomy assessment scale (LMAS). Finally, to ensure reliability and validity of the tool, we used the LMAS to assess the video recordings of eight expert and eight novice performance of LM. We were able to demonstrate that the LMAS is a valid and reliable tool for assessing performance of LM surgery.

PREFACE

The need for rigorous assessment of surgical skills in neurosurgery

Spinal surgery is one of the fastest growing areas of medicine. It is rapidly evolving with new technology emerging on a yearly basis. Spinal surgery has always been associated with complications, some of which are serious enough to cause patients permanent disability. Indemnity premiums for spinal surgery have skyrocketed as there has been increased litigation related to the complications of spinal surgery. More recently public examples of the devastating impact of poorly trained spine surgery have been illustrated through the book and then television representation of Dr Christopher Duntsch who devastated spinal patients in Texas. Christopher Duntsch was a spinal surgeon trained at the reputable Semmes-Murphey Clinic in Memphis Tennessee. He was thought to have participated in fewer than 200 spinal surgical procedures. He was able to work in Texas and unfortunately due to his lack of training maimed and killed several patients[1]. What was evident was the lack of training that Dr Duntsch had. This and other examples in other forms of surgery have led to rigorous credentialling processes for spinal surgeons. Medicine in the modern era and especially spinal surgery has moved towards a competency-based model of training. However, the means with which we assess spinal surgery trainees still utilises traditional Halstedian rubrics.

Development and validity of assessment tools

The development of assessment tools is a careful process guided by a number of educational theories. The first step is to clearly define the purpose and the specific skills or knowledge to be measured, a task typically informed by research. This foundation then dictates the type of assessment method and how its individual items are created. The entire process is built on

two pillars: validity, which ensures the tool accurately measures its intended target, and reliability, which guarantees consistent results over time. After the initial design, the tool undergoes a rigorous process of pilot testing, statistical analysis, and iterative revisions based on feedback from both experts and users to ensure it is fair, useful, and contextually appropriate[2-4].

Procedure specific assessment tools have been developed in other surgical fields including general surgery and gynaecology[5-7]. These tools have been demonstrated to improve patient outcomes and reduce complications[2,8]. It was in this context that the idea for this thesis was borne.

Education theory and conceptual model

The theoretical foundation for this research is post-positivist pragmatism, a philosophy that guided the development of the LMAS assessment tool[9]. This approach aims to achieve objective assessment in surgical learning by acknowledging that complete objectivity is impossible. Instead, it seeks to get as close as possible by emphasising practical utility, reproducibility, and evidence-based methods[10]. The aim is to develop the “probable” truth. This differs from the “positivist” philosophy of knowledge that has historically dominated medical education, where “objectivity” exists[10].

This perspective is particularly effective for surgical training because it values reliability and fairness. Unlike traditional assessment methods that might ignore the context of a training environment, post-positivist pragmatism integrates these real-world influences into the design of the LMAS. The use of standardised criteria, for example, reduces subjective

interpretation and ensures consistent, scientifically sound, and actionable feedback for trainees. This approach was essential for the thesis's goal of creating a dependable and context-sensitive assessment for lumbar microdiscectomy.

Work-based assessment as a concept is relatively novel. There has been an emergence in the medical education literature, that provision of feedback, within a framework of expected competencies, allows a trainee to be steered towards achieving a desired outcome[11,12]. It is from this perspective that I chose to develop a procedure-specific learning tool for the index operation of spinal surgery, lumbar microdiscectomy. Feedback has been demonstrated to promote learning in students by informing trainees of progress, inform them of learning needs and motivating trainees[11]. Formative assessment using feedback to achieve goals of learning has been demonstrated to increase time spent preparing for skills training sessions. In surgical training programs, work-based assessment has been used to improve trainee achievement of technical and non-technical competency. However, to date, procedure-specific assessment tools have been rare, and none have been developed for lumbar microdiscectomy.

Thesis structure and aims

I therefore sought to identify what educational tools currently existed within the literature for spinal surgery trainees. The general outline of the thesis is seen in Figure 1 below.

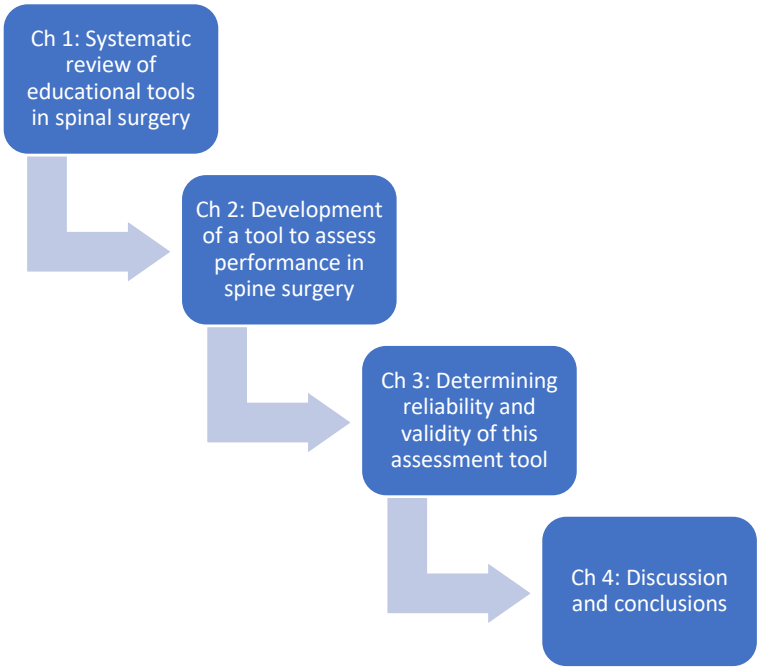


Figure 1: Thesis flowchart

Chapter 1 is the systematic review that was performed to identify if there were existing procedure-based assessment tools. Once I identified that none had been developed, I set about the task of developing a procedure-based tool specific for lumbar microdiscectomy. Using a hierarchical task analysis and Delphi methodology I was able to develop the Lumbar Microdiscectomy Assessment Scale (LMAS) in Chapter 2 with the input of a number of experienced spinal surgeons from multiple institutions. Once this was developed it was important that it was validated to confirm that it was a reliable, feasible tool for assessing performance of lumbar microdiscectomy. As described in Chapter 3 I did this using video analysis of expert and novice spinal surgeons. Having confirmed that the LMAS is valid and reliable, in Chapter 4 I discuss future steps and some concluding thoughts.

While these are the very first steps in developing and validating a tool for assessing performance of lumbar microdiscectomy surgery, it is very exciting in the context of reforming competency-based medical education within spinal surgery. My hope is that I can validate the LMAS at other institutions and obtain buy-in from supervisors, trainees and spinal surgery training societies. In this manner, change the way we train surgeons to perform lumbar microdiscectomy, and then other spinal operations. It is only by training our future surgeons in the best possible way, that we can prevent the catastrophe of Dr Duntsch from happening again.

CHAPTER 1: SYSTEMATIC REVIEW OF EDUCATION TOOLS IN SPINE SURGERY

In this chapter we describe a systematic literature review of educational tools used for spinal surgery trainees.

This chapter is in manuscript form with view to submission in a peer reviewed journal.

ABSTRACT

Purpose

Spinal surgery is an ever increasing around the globe. The potential for serious complications exists and there are examples of surgeons with poor training causing devastating patient outcomes such as the infamous Dr Duntsch. Therefore, we aimed to understand the existing educational tools that are being utilised in spinal surgery and performed a systematic review to do this.

Methodology

Using Medline, Embase and Cochrane review databases we conducted a search for: surgery, spine, education, training, assessment, virtual-reality, simulation, laminectomy, pedicle screw

Results

The systematic search after exclusion criteria identified 30 articles on spinal surgical education with trainee involvement. The educational tools involved included saw-bone models, 3D printed models, custom designed simulation models, cadaveric models, virtual reality simulation and augmented reality tools. There were no procedure-based assessment tools used to evaluate performance during simulated or actual spine surgery.

Conclusion

We performed a systematic review of spinal surgical educational tools involving trainees. We identified that whilst there were several educations tools used to simulate spine surgery, there were no procedure-based assessment tools in spinal surgery. Through this systematic review a gap in the literature has been identified and provided the impetus to develop a procedure-based tool for use in assessment of lumbar microdiscectomy, the most common procedure in spine surgery.

INTRODUCTION

Surgical training has traditionally been an apprenticeship model first introduced by Halsted in the 19th century. The adage “see one, do one, teach one” has long purveyed medicine and especially surgery. In the past trainees have therefore learned and practised surgery simultaneously[13,14]. However, as we continue on in the 21st century, medicine and surgery have evolved, patient and societal expectations have altered, work-hour regulations have changed and instead of a longitudinal assessment of experiential learning, competency assessment is becoming the standard of surgical training. Competency-based training sets predefined standards, technical and non-technical, that trainees must achieve by completion of training[15]. Despite moving towards competency-based assessment, traditional assessment tools such as objective structured clinical examinations (OSCEs) or viva examination remain prevalent[2]. There remains a focus on case logbook numbers in assessment which indicate the volume of surgery, not how well the surgery has been done. These assessment tools provide only a snapshot of a trainee’s performance and are therefore not necessarily an accurate reflection of competence.

Training is evolving for the modern era. Virtual reality 3D simulation training tools, have become part and parcel of training[16-18]. However, as competency-based training establishes itself as the leading training model, they must include tools to assess competence. These tools should be rigorously developed, and ideally be standardised, valid, and reproducible in a global context[15]. The earliest competency-based assessment tools were global assessment scores, the Objective Structured Assessment of Technical Skill (OSATS)[19] and Global Rating Scale (GRS)[20]. These tools have formed the basis for a number of more specific procedure-

based assessment tools[21]. Global rating scales can evaluate generic surgical skills by breaking down operative competence into smaller, component skills. Each skill can be rated on a Likert scale. The Objective Structured Assessment of Technical Skill (OSATS) is the most researched observational assessment tool[19]. Introduced in 1997, it takes a comprehensive approach to assessing technical skills. OSATS focuses on six universal surgical skills, such as tissue handling and surgical flow. Each skill is rated on a Likert scale to generate an overall performance score. OSATS has proven reliable, feasible, and valid, with adaptations for various medical fields. In response to the rise of laparoscopic surgery, OSATS has been modified for this specific technique. A study by Aggarwal et al. investigated the reliability and validity of both original and modified OSATS tools in assessing laparoscopic cholecystectomy performance[22]. The original OSATS tool demonstrated moderate reliability, with both inter-rater and test-retest reliability scores with a Cronbach alpha of 0.72. The modified OSATS tool showed slightly higher inter-rater reliability (0.76) but similar test-retest reliability (0.72). Both tools successfully distinguished between experienced and inexperienced surgeons, indicating their validity[22].

The Global Assessment of Laparoscopic Skill (GOALS) tool, developed in 2005, assesses key laparoscopic skills using a five-item Likert scale[23]. It has shown high inter-rater reliability (0.89) and has been validated through video-based performance assessments. A significant study by Birkmeyer et al. linked lower OSATS scores in bariatric surgery to poorer patient outcomes, including higher complication and mortality rates, longer procedure times, and increased reoperation rates[24]. While beneficial, global rating scales do not provide procedure-specific feedback to the trainee[25].

Procedure-based assessment tools have the advantage over global rating scales in that they can be validated for a specific procedure by experts in that procedure, and are reproducible across varying contexts[5,7,25,26]. These tools represent a hybrid between global rating scales and checklists, which are more prescriptive tools[7,25-27]. Checklists break operations down into individual steps and require surgeons to complete them in a specific order. Deviating from this sequence can lead to a low checklist score. While these checklists can offer structured feedback to trainees, their rigid format has been linked to issues with validity and reliability[7,25,27]. Procedure-specific tools are less rigid but they still provide specific formative feedback to the trainees. These tools are feasible and validated for each of the operations they are designed to assess. When used in a standardised curriculum for a specific operation in a training setting, surgical trainees that used the objective assessment tool had a superior improvement in knowledge and technical skills than those that did not[6]. While procedure-specific tools are known to provide an important means of assessing a surgeon's performance in a specific procedure, they are limited to a few laparoscopic surgical procedures at present[7,25,26,28-30].

Spine surgery is one of the fastest growing areas of surgical practice in the Western world[31]. Complications associated with spine surgery can have significant impacts on the quality of life of patients. Therefore, spine surgery is also one of the most litigated specialties with increasing malpractice claims resulting in increasing indemnity insurance for spine surgeons[32,33]. The troubling case of Dr Christopher Duntsch outlines all too clearly what can happen if a poorly trained spine surgeon enters into practise and performs routine spine

surgery[1]. Questions remain around his training, his low case numbers and supervision. His is obviously a rare and hopefully singular exception. However, even well-trained surgeons performing complex spine surgery can have one documented complication in up to 87% of cases[1,34]. Even for a simple spine operation, a lumbar microdiscectomy, complications occur at an increased level in a public hospital with neurosurgical trainees than in the private setting with neurosurgical consultants[35]. Public hospital patients have longer length of stay, longer duration of surgery, increased rate of recurrent surgery than consultant surgeons working in private practice[35,36]. Therefore, it is important that neurosurgical and orthopaedic trainees learning spine surgery are provided with appropriate training to safely perform this common operation.

Traditionally, surgical training was a time-based approach, requiring a certain number of years of experience. However, the focus has shifted to a competency-based model. The Royal Australasian College of Surgeons uses a competency-based model of training across all its subspecialties including neurosurgery and orthopaedic surgery. However, there are still minimum times to completion of these competencies. Therefore, there is still a time-based longitudinal approach to surgical training. To address concerns about the effectiveness of this shift, there has been a push to replicate real-life surgical experiences through simulation[6,37]. There are various simulation models used in surgical training, including saw bone models, 3D printed models, custom designed models, cadaveric models, and virtual reality models.

Despite the sophistication of these models, current assessment methods in spinal surgery are not procedure-specific. As discussed above procedure-specific assessment tools improve performance of trainees. Therefore, in line with the PRISMA statement we performed a systematic review of tools currently being used in education in spinal surgery[38]. We aimed to identify the tools that exist, and if any of these were procedure-specific assessment tools.

METHODS

On February 6th 2022, Medline, Embase and Cochrane review databases were searched. The following search terms were used to conduct the study: surgery, spine, education, training, assessment, virtual-reality, simulation, laminectomy, pedicle screw (Figure 1.1). MeSH terms were exploded for relevant terms that met the inclusion criteria, and Boolean terms were used to combine search terms. The searches were limited to English-language, peer-reviewed articles published between 2002 and February 6, 2022.

Search n	Query	Sort By	Filters	Search D	Results	Time
1	surg*			"surg*"[A	#######	7:10:42
2	residen*			"residen*	368,441	7:11:20
3	trainee*			"trainee*"	34,566	7:11:39
4	(residen*) OR (trainee*)			"residen*	392,562	7:12:26
5	(surg*) AND ((residen*) OR (trainee*))			"surg*"[A	58,375	7:12:55
7	objective assessment			("goals"[I	939,386	7:14:20
8	global assessment			("global"[.	200,408	7:14:29
9	procedure based assessment			("method	541,588	7:14:42
10	OSATS			"OSATS"	389	7:14:52
11	((((assessment) OR (objective assessment)) OR (global assessment)) OR (procedure based assessment)) OR (OSATS)			"assess"	#######	7:15:21
12	rating			"rated"[A	325,044	7:15:36
13	tool			"tool"[All	585,312	7:15:43
14	scale			"scale s"[	#######	7:15:48
17	(((surg*) AND ((residen*) OR (trainee*))) AND (((((assessment) OR (objective assessment)) OR (global assessment)) OR (procedure based assessment)) OR (OSATS))) AND ((((rating) OR (tool)) OR (scale)) OR (measure))			"surg*"[A	8,075	7:16:39
18	(((surg*) AND ((residen*) OR (trainee*))) AND (((((assessment) OR (objective assessment)) OR (global assessment)) OR (procedure based assessment)) OR (OSATS))) AND ((((rating) OR (tool)) OR (scale)) OR (measure))		from 200	("surg*"[A	7,743	7:17:12
19	(((surg*) AND ((residen*) OR (trainee*))) AND (((((assessment) OR (objective assessment)) OR (global assessment)) OR (procedure based assessment)) OR (OSATS))) AND ((((rating) OR (tool)) OR (scale)) OR (measure))		Full text,	("surg*"[A	7,513	7:18:38
20	(((surg*) AND ((residen*) OR (trainee*))) AND (((((assessment) OR (objective assessment)) OR (global assessment)) OR (procedure based assessment)) OR (OSATS))) AND ((((rating) OR (tool)) OR (scale)) OR (measure))		Full text,	("surg*"[A	6,434	7:19:46
21	(((surg*) AND ((residen*) OR (trainee*))) AND (((((assessment) OR (objective assessment)) OR (global assessment)) OR (procedure based assessment)) OR (OSATS))) AND ((((rating) OR (tool)) OR (scale)) OR (measure))		Full text, I	("surg*"[A	6,382	7:27:10
22	((rating) OR (tool)) OR (scale)		Full text, I	("rated"[A	#######	7:30:36
23	(((surg*) AND ((residen*) OR (trainee*))) AND (((rating) OR (tool)) OR (scale) AND ((fft[Filter]) AND (english[Filter])))) AND (((((assessment) OR (objective assessment)) OR (global assessment)) OR (procedure based assessment)) OR (OSATS))		Full text, I	("surg*"[A	4,435	7:31:51
24	(((surg*) AND ((residen*) OR (trainee*))) AND (((rating) OR (tool)) OR (scale) AND ((fft[Filter]) AND (english[Filter])))) AND (((((assessment) OR (objective assessment)) OR (global assessment)) OR (procedure based assessment)) OR (OSATS))		Full text, I	("surg*"[A	4,344	7:32:10
25	(((objective assessment) OR (global assessment)) OR (procedure based assessment)) OR (OSATS)		Full text, I	((("goals"	#######	7:32:51
26	(((surg*) AND ((residen*) OR (trainee*))) AND (((rating) OR (tool)) OR (scale) AND ((fft[Filter]) AND (english[Filter])))) AND ((((objective assessment) OR (global assessment)) OR (procedure based assessment)) OR (OSATS) AND ((fft[Filter]) AND (english[Filter])))		Full text, I	("surg*"[A	3,173	7:33:28
27	(((surg*) AND ((residen*) OR (trainee*))) AND (((rating) OR (tool)) OR (scale) AND ((fft[Filter]) AND (english[Filter])))) AND ((((objective assessment) OR (global assessment)) OR (procedure based assessment)) OR (OSATS) AND ((fft[Filter]) AND (english[Filter])))		Full text, I	("surg*"[A	3,087	7:33:41
28	(((surg*) AND ((residen*) OR (trainee*))) AND (((rating) OR (tool)) OR (scale) AND ((fft[Filter]) AND (english[Filter])))) AND ((((objective assessment) OR (global assessment)) OR (procedure based assessment)) OR (OSATS) AND ((fft[Filter]) AND (english[Filter])))		Full text, I	("surg*"[A	2,556	7:33:59
29	((objective assessment) OR (procedure based assessment)) OR (OSATS)		Full text, I	((("goals"	#######	7:47:40
30	((((objective assessment) OR (procedure based assessment)) OR (OSATS) AND ((fft[Filter]) AND (english[Filter]))) AND (((rating) OR (tool)) OR (scale) AND ((fft[Filter]) AND (english[Filter])))) AND ((surg*) AND ((residen*) OR (trainee*)))		Full text, I	((((("goals	3,020	7:48:13
31	((((objective assessment) OR (procedure based assessment)) OR (OSATS) AND ((fft[Filter]) AND (english[Filter]))) AND (((rating) OR (tool)) OR (scale) AND ((fft[Filter]) AND (english[Filter])))) AND ((surg*) AND ((residen*) OR (trainee*)))		Full text, I	((((("goals	2,444	7:50:01
32	((((objective assessment) OR (procedure based assessment)) OR (OSATS) AND ((fft[Filter]) AND (english[Filter]))) AND (((rating) OR (tool)) OR (scale) AND ((fft[Filter]) AND (english[Filter])))) AND ((surg*) AND ((residen*) OR (trainee*)))		Full text, /	((((("goals	1,566	7:48:37
33	rating scale		Full text, I	(("rated"[	165,451	8:04:40
34	(tool) OR (rating scale AND ((fft[Filter]) AND (english[Filter])))		Full text, I	("tool"[All	683,185	8:05:19
35	(((((objective assessment) OR (global assessment)) OR (procedure based assessment)) OR (OSATS) AND ((fft[Filter]) AND (english[Filter]))) AND ((tool) OR (rating scale AND ((fft[Filter]) AND (english[Filter]))) AND ((fft[Filter]) AND (english[Filter])))) AND ((surg*) AND ((residen*) OR (trainee*)))		Full text, I	((((("goals	1,935	8:06:32
36	surgical performance		Full text, I	(("surgica	683,658	8:07:19
37	technical skill		Full text, I	(("technic	11,452	8:07:31
38	(surgical performance AND ((fft[Filter]) AND (english[Filter]))) OR (technical skill AND ((fft[Filter]) AND (english[Filter])))		Full text, I	((("surgic	692,222	8:07:47
39	((surgical performance AND ((fft[Filter]) AND (english[Filter]))) OR (technical skill AND ((fft[Filter]) AND (english[Filter]))) AND ((fft[Filter]) AND (english[Filter]))) AND ((((((objective assessment) OR (global assessment)) OR (procedure based assessment)) OR (OSATS) AND ((fft[Filter]) AND (englis		Full text, I	((((("surgi	1,226	8:10:02
40	((surgical performance AND ((fft[Filter]) AND (english[Filter]))) OR (technical skill AND ((fft[Filter]) AND (english[Filter]))) AND ((fft[Filter]) AND (english[Filter]))) AND ((((((objective assessment) OR (global assessment)) OR (procedure based assessment)) OR (OSATS) AND ((fft[Filter]) AND (englis		Full text, I	((((("surgi	960	8:10:10
41	((surgical performance AND ((fft[Filter]) AND (english[Filter]))) OR (technical skill AND ((fft[Filter]) AND (english[Filter]))) AND ((fft[Filter]) AND (english[Filter]))) AND ((((((objective assessment) OR (global assessment)) OR (procedure based assessment)) OR (OSATS) AND ((fft[Filter]) AND (englis		Full text, I	((((("surgi	993	8:10:46

Figure 2.1: Search terms used in this review

To ensure the inclusion of relevant studies, we adopted a rigorous inclusion and exclusion criteria. Articles were included if they specifically focused on the education of adult spine surgery trainees. This criterion excluded studies that did not involve trainees or fellows, such as those solely focused on the development of educational models without implementation. Additionally, articles that were primarily commentaries, reviews, or abstracts were excluded to maintain a focus on empirical research. By adhering to these criteria, we aimed to identify studies that directly address the educational needs of spine surgery trainees and provide valuable insights into effective training strategies.

Three independent reviewers performed a title and abstract screening followed by a full text review. Initially, eligibility was deemed by evaluating study titles and abstracts. Those studies that were not relevant were excluded. If it was unclear whether the study should be included from the title and abstract alone, the full text was assessed. Studies which involved education and adult spine surgery were included. Duplicates and review articles were excluded once bibliographies were searched.

Any disagreements were resolved by consensus discussion amongst all three reviewers with any discrepancies decided upon by the first author. The following exclusion criteria was used during the full text screening process:

- Not focused on training or education
- Review article
- Not adult spine surgery

- Did not involve trainees
- Was not a full text article

A data acquisition form was used which included study design, participants and type of education. Due to the breadth of the review where multiple education types across spine surgery were being reviewed, a meta-analysis could not be performed. Therefore, we grouped the studies by education type to provide an overview of the education tools that are currently in use in spine surgery.

RESULTS

Study Selection

The systematic search revealed 8,863 articles from Medline, Pubmed and Embase. Of these there were 1,563 duplicates which were excluded. 7,300 title and abstracts were screened to include 70 articles for full text review. 8 review articles which did not present new research were excluded. Of these 27 were included and from hand-searching the references of these articles a further 3 were included for a total of 30 articles (Figure 1.2).

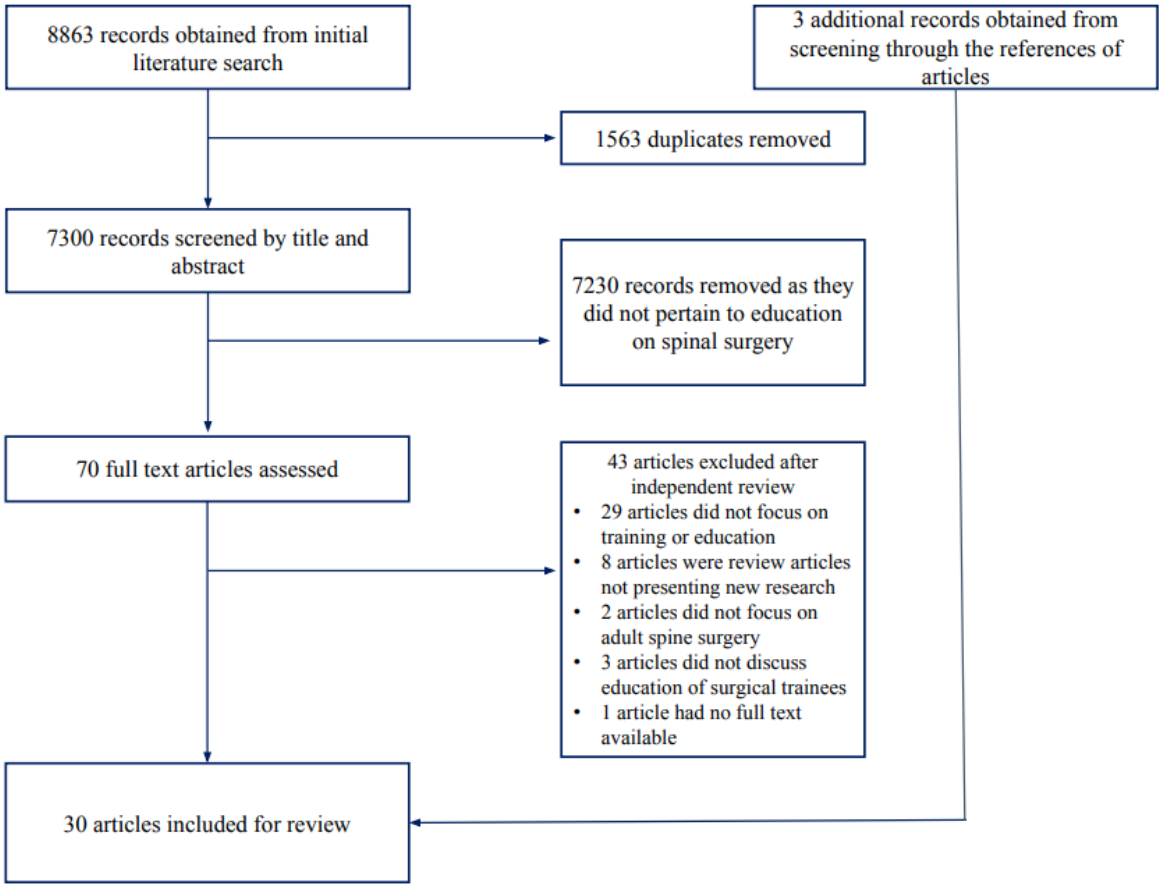


Figure 3.2: PRISMA Flowchart[38]

Study characteristics- demographics/specialty

Thirty studies were included which were published between 2007 and 2021. 20 out of 31 articles were published in the USA. A number of different educational designs were used including custom designed simulation models, sawbone models, 3D printed models, cadaver simulation, virtual reality simulation, augmented reality simulation and computer assisted navigation simulation (Table 1.1). Both neurosurgery and orthopaedic surgery were involved in the included studies. All included studies had trainee involvement in the use of the proposed educational tool. The number of trainees involved ranged from 3 up to 72.

Objective assessment of performance in lumbar microdiscectomy

Author	Year	Country	Design	Participants	Assessment timeframe	Methodology	Procedure-based assessment tools used	Measured outcomes	Cost	Conclusion	Type of validity
Andermann et al.,	2014	Germany	Custom designed model simulation	4 novice surgeons (not specified PGY level)	One-day workshop		OSATS	Improved intra-operative orientation, improved surgical performance	Not specified	The designed model helped trainees to improve their confidence in spinal surgery	Face
Bhatia et al.,	2020	USA	Custom designed simulation model	72 neurosurgery orthopaedic residents	Same day assessment		Assessment of correct placement and identification of breach	Accuracy of pedicle screw trajectory and ability to detect pedicle breach	$30	The model helped trainees to identify pedicle wall breaches	Construct
Boody et al.,	2017a	USA	Sawbones model	19 orthopaedic residents and medical students	Same day assessment	Intervention group provided 20 minutes to work with the models compared to	OSATS, PPDIS Assessment of correct placement and identification of breaches	Improvement in OSATS score, number of breaches, and number of breaches identified	$72	The intervention group performed significantly better than the control group in both placement of screws and identification of breaches	No discussion on validation, attempt at transfer
Boody et al.,	2017b	USA	Sawbones model	20 orthopaedic residents and medical students	Same day assessment	Intervention group provided 40-minute teaching	OSATS, PPDIS, ODS	Assessment in completion of decompression of 4 regions	$240	The intervention group performed significantly better in all domains	Not mentioned, attempt at transfer

						session by senior resident					
Bow et al.,	2020	USA	3D printed model	6 neurosurgery residents	Same day assessment	All participants underwent completion of three modules surrounding S2AI screw placement	N/A	Assessment of accuracy of S2AI screw placement, direction of breach	$17	Following training on the module, trainees performed better on the first two modules (3rd module was identification of breach)	Quasi-transfer
Buchanan et al.,	2019	USA	Cadaver simulation	8 neurosurgery trainees	2-week interval	All participants underwent assessment of dural leak repair	Likert scale on confidence in completing dural leak repair	Time taken for dural tear closure, confidence in ability to perform a repair	Not specified	Following training, trainees were faster at closing dura with lower rates of persistent leak	Quasi-face
Calio et al.,	2017	USA	Cadaver simulation	17 orthopaedic residents	Same day assessment	Junior residents opened and closed a spinal wound, while senior residents performed a decompression	Likert scale on confidence and safety of performing the allocated steps	Assessment on the confidence in performing the steps allocated during training	Not specified	Residents at all levels felt more confident and safe in performing the allocated steps	Quasi-face
Chen et al.,	2021	Canada	Virtual reality simulation	28 neurosurgery and orthopaedic surgery	Same day assessment	All participants performed a single level spinal decompression	N/A	Assessment in the knowledge of anatomy and pathophysiology of the surgery	Not specified	Most participants felt it improved their knowledge	Transfer, construct

				residents and fellows							
Chitale et al.,	2013	USA	Simulation model	8 neurosurgery trainees	Same day assessment	All participants underwent a practical and written tutorial lasting 2 hours on pedicle screw placement	Assessment on accuracy of screw placement	Accuracy of screw placement, time using fluoroscopy	Not specified	Whilst the result trended towards significant, there was no conclusive improvement	N/A
Delgado-Fernandez et al.,	2021	Spain	Sawbones	3 neurosurgery trainees	Same day assessment	All participants performed simulated pedicle screw placement	OSATS, PPDI, accuracy of screw placement	Accuracy of screw placement using post-operative scans	Not specified	OSATS and PPDI scores significantly improved following simulation experience	Face
Ferguson et al.,	2019	UK	Simulation model	10 neurosurgery trainees	Same day assessment	All participants received 15 minutes practice after their first attempt at dural closure	OSATS	Time taken for dural closure, quality of dural closure	Not specified	Significant improvement noted in trainees on both time of closure and quality of closure	Face
Gardeck et al.,	2020	USA	Simulation model	15 neurosurgery and orthopaedic surgery trainees	10-day interval	All participants underwent training with one practice session then an additional assessment	PPDIS, OSATS, accuracy of screw placement and timing	Accuracy of screw placement and time to place the screw	Not specified	Both PPDIS and OSTAS scores increased over time, but no change noted in the accuracy of placement	Transfer

Ghobrial et al.,	2013	USA	Sawbones	4 neurosurgery trainees, 1 consultant, 1 PA	Same day assessment	All participants had at least one attempt at dural closure	Ad-hoc complication of questions	Speed and quality of closure	Not specified	Amongst the trainees, there was uniform improvement on the scores with improved time and quality of closure	Content
Ghobrial et al.,	2015	USA	Sawbones	20 neurosurgery trainees	Pre-test done 48hrs before the assessment	All participants had three attempts at dural closure and one attempt at cervical foraminotomy	OSATS	OSAT assessment of various components of dural closure and foraminotomy	Not specified	Improvement in both dural closure and cervical foraminotomy	Face
Gottschalk et al.,	2015	USA	Cadaver or sawbones simulation	15 orthopaedic trainees	Same day assessment	Participants divided into three groups, one control group without practice, one group with 3D simulation practice on sawbones model, and one cadaver group with 3D simulation practice	N/A	Assessment of the accuracy of screw placement	Total cost $12,000	Both the sawbones and cadaver group significantly improved compared with the control group	Transfer
Harrop et al.,	2013	USA	Custom made	8 neurosurgery	Same day assessment	Participants underwent a	Custom assessment	Accuracy and technical	Not specified	Participants demonstrated	Transfer

			simulation model	trainees and 1 specialist		2-hour teaching component with both lectures and a didactic component	based off OSATS	proficiency in screw placement		significant improvement following simulation	
Hou et al.,	2018a	China	Virtual reality	10 neurosurgery trainees	Same day assessment	Participants divided into control and simulation group. Simulation group given 30 minutes to practice on the simulator placing pedicle screws, control group received didactic education	N/A	Accuracy of screw placement	Not specified	The simulation group demonstrated improved accuracy in placement of pedicle screws	Transfer
Hou et al.,	2018b	China	Virtual reality	10 orthopaedic trainees	Same day assessment	Participants divided into control and simulation groups. Simulation group provided 30 minutes of practice on virtual reality	N/A	Accuracy of screw placement	Not specified	The simulation group had significantly higher rates of accurate screw placement than the control group	Transfer

						simulation, control group given didactic feedback					
Luca et al.,	2020	Italy	Virtual reality	7 orthopaedic trainees and 3 surgeons	Same day assessment	Participants completed the VR teaching modules designed for this study	N/A	Evidence of any major operative errors	Not specified	For the trainees, use of the simulation significantly reduced the rate of simulated operative errors on the assessment	Transfer
Luciano et al.,	2011	USA	Augmented reality	51 neurosurgery fellows and trainees	Same day assessment	Participants completed a 5 minute training module then underwent a testing period	N/A	Accuracy of screw placement	Not specified	For trainees using the software there was improved accuracy in pedicle screw placement	Transfer
Luciano et al.,*	2013	USA	Augmented reality	63 neurosurgery fellows and trainees	Same day assessment	As above	N/A	Accuracy of screw placement	Not specified	Results between experiments were similar between the pre-test and post-test	Transfer
Mendelis et al.,	2020	USA	Simulation model	6 inexperienced trainees	Same day assessment	Trainees were given 2 screws to place, then underwent education with subsequent assessment on	Ad-hoc compilation of scoring	Accuracy of screw placement, ability to identify breach	Not specified	Significant improvement in screw placement after education session	Transfer

						2 further pedicle screws					
Park et al.,	2018	South Korea	3D printed simulation model	2 orthopaedic trainees	Same day assessment	Trainees inserted 100 pedicle screws total, monitoring for the accuracy of each screw	N/A	Accuracy of screw placement	Not specified	With further experience working with the model spines, accuracy of screw placement improved	Transfer
Podolsky et al.,	2010	Canada	3D printed simulation model	37 neurosurgery and orthopaedic trainees	Same day assessment	Participants allocated to control or simulation groups. Both groups received 3 hours teaching, simulation group had an additional 1 hour with the simulation model to practice before the assessment.	N/A	Accuracy of screw placement	Not specified	There was no difference in the accuracy of screw placement between the simulation and control groups	Face
Rambani et al.,	2014	UK	Computer assisted navigation simulation	12 orthopaedic trainees	Not specified	Participants split into control and simulation groups, unspecified	Ad-hoc compilation of questions designed for the study	Accuracy of screw placement	Not specified	There was overall improvement in pedicle screw placement for	Transfer

						amount of time spent training before assessment of placement				the simulation group	
Ray et al.,	2013	USA	Simulation model	3 neurosurgery trainees, 2 medical students and 1 consultant	Same day assessment	Participants were given a total of 90 minutes for education including simulation experience then testing	Customs designed OSATS	Assessment was done based on accuracy of each step of an ACDF	Not specified	There was a significant improvement in testing scores amongst all trainees and medical students who underwent the simulation training	Face
Sundar et al.,	2016	USA	Sawbones model	8 neurosurgery trainees and 2 medical students	Assessment 1-2 weeks post training	Participants were divided into a control and simulation group, with both groups given a 1hr lecture on the basics. The study group experienced navigation-based teaching on a sawbones model	N/A	Accuracy of pedicle screw placement	Not specified	The study group had significantly less suboptimal screw placements in cervical, thoracic and lumbar models.	Face
Tanner et al.,	2017	USA	Sawbones	20 orthopaedic trainees	Same day assessment	Participants were randomly allocated	Global rating scale	Accuracy of screw placement and	Not specified	Senior residents were faster at screw placement	Construct

						either right or left side. Both groups placed pedicle screws from T4 to L4		total operative time		and had less surgical errors than junior residents	
Tortolani et al.,	2013	USA	Cadaver	3 orthopaedic trainees	Assessment done one month after teaching	Participants were given a 1hr instructional session by the senior author. All residents were given cadavers to practice screw placement. One month later, the same process was repeated	Ad-hoc compilation designed for the study	Accuracy of pedicle screw placement	Not specified	During the second session, the trainees demonstrated a significant improvement in their accuracy of screw placement	Face
Walker et al.,	2009	USA	Simulation model	8 neurosurgery trainees	Same day assessment	Participants were given an activity based on their level in training and given practice time with the simulator. They were asked to rate their confidence	Ad-hoc self-assessment questionnaire	Trainee confidence in performing MIS procedures	Not specified	Trainees had significant improvement in confidence	Face

						following the training					
Woodrow et al.,	2007	Canada	Synthetic sawbones model	12 neurosurgery trainees	Same day assessment	Participants were given a model to place 10 pedicle screws	N/A	Accuracy of screw placement	Not specified	Consultant surgeons performed better with the simulated model than residents. Residents had improved confidence after the assessment	Construct

Table 1.1: Data extraction table from systematic review

Type of educational tool

Sawbone model

Sawbone models are a commonly used simulation tool for spine education due to their ability to replicate the complex 3D anatomy of the spine whilst avoiding the need for sourcing cadaveric or animal bone. They are made from foam and other plastic composites to achieve specific characteristics depending on the purpose of the model. Sawbones are also cheaper than cadaveric specimens. Sawbones however cannot replicate the internal architecture of real bone and its viscoelastic properties. Eight of the identified studies in this review used sawbone models with the majority testing screw placement[17,39-43] with one assessing lumbar laminectomy[44] and one assessing cervical foraminotomy[45]. The majority of the studies used sawbone models as an assessment tool to evaluate performance of the primary intervention. Boody et al used a bioskills training module where groups of trainees were randomised to either an interactive presentation or reading standard texts of their own choice[39]. The bioskills training module group had an improvement in OSATS as well as pedicle screw placement and identifying their own breaches and in their second study that assessed lumbar laminectomy, better decompressions were performed[39,44]. In their study of cortical bone trajectory screws Delgado-Fernandez et al demonstrated improvement in OSATS and Physician Performance Diagnostic Inventory Scale (PPDIS), which is a five-point self-rating score, for trainees using the sawbone model with 3D fluoroscopic navigation[17]. In a group of orthopaedic trainees, Gottschalk et al used sawbones and cadaver models for cervical lateral mass screw placement education[42]. 3D navigation was used for training of the correct trajectory and found that with training, angle of screw placement, facet violations and nerve injuries all improved. They also demonstrated no significant differences between

sawbone models and cadavers in educational and training value[42]. In a scoliosis sawbone model Tanner et al found that screw placement was performed quicker by senior trainees, however, screw placement accuracy was similar despite prior verbal instruction[41].

3D printed model

3D printing is a rapidly growing technology that has many applications in medical education[46]. Two of the included studies used 3D printing technology for their training simulation models[47,48]. The 3D printed pelvis made by Bow et al was used as an assessment tool for S2-Alar-Iliac screw placement via C-arm fluoroscopy. Their educational tool improved both screw placement and obtaining C-arm fluoroscopic images to identify breaches[47]. Park et al used a 3D printed spine model as a training aid for pedicle screw insertion, and demonstrated an improvement in pedicle screw placement[48]. The 3D polymers used for printing with these studies were combinations of polypropylene and polylactic acid / polyvinyl alcohol filaments.

Custom designed simulation model

A range of custom designed models were identified in this review, aiming to replicate the optic and haptic feedback of real surgery by using a variety of mediums including wood, paper, cardboard, plastic, latex, silicone and other synthetic materials. A common theme to the development of the majority of the models was to create a low-cost substitute that would be an accurate representation of the surgical environment.

A custom designed model for simulation was used in 11 of the reviewed studies, assessing various procedures including lumbar discectomy[49], screw placement[18,50-53], dural closure[45,54,55], laminoforaminotomy[56], Anterior Cervical Discectomy and Fusion (ACDF)[52] as well as a variety of Minimally Invasive Spine (MIS) procedures [53,57]. Adermann et al developed a custom-made spine model, which replicated all the anatomical features of an L4/5 disc[49]. Cerebrospinal Fluid (CSF) and dura were also replicated. Pressure sensors and ultrasonic signals were used to simulate unintended durotomy and compression forces on the nerve root. Following validation of the model, trainees underwent a workshop where they were able to achieve 100% of their predefined goals and improvement in their self-assessment of discectomy[49]. The Fundamental of Spine Surgery simulator, a custom designed model by Bhatia et al, cost $30 USD to make and was used by 72 orthopaedic and neurosurgery trainees[50]. The vast majority (89%) of trainees who used the simulator felt it would improve their pedicle screw placement. Gardeck et al used the Medtronic TruTrainer (Medtronic Sofamor Danek, Inc., Memphis, TN, USA) in conjunction with 3D navigation for thoracolumbar pedicle screw placement training[51]. Their trainees demonstrated an improvement in OSATS and screw placement time, although accuracy and angulation error were no different[51]. A low-cost pedicle screw training model was developed by Mendelis and collaborators which demonstrated that with instruction and training, pedicle screw accuracy, time for screw placement, and screw angulation improved[18]. Ghobrial et al developed a dural repair and cervical laminoforaminotomy simulation model[45,54,56]. Using a sawbone and then customising a dural substitute to mimic dura they were able to use the custom model to simulate dural repair. They provided instruction prior to using the model and demonstrated improvement in time to repair, leak rate and technical improvements overall[45,54]. The laminoforaminotomy model was developed in conjunction with Stryker

(Kalamazoo, Michigan), trainees were administered a pretest on technical performance of a left C7/T1 laminoforaminotomy[56]. They then had an instructional lecture followed by time on a simulation model. Finally, they had a post-test using the same questions as the pre-test, however, reorganised randomly. All trainees demonstrated improvement on their custom assessment tool[56]. Following initial testing with these two simulators, a neurosurgical curriculum was developed which again found significant improvement in performance on the post-test assessment as above[45]. The other dural closure model developed by Ferguson et al involved a tubular roll of wallpaper, a tubular glove-roll complex and a wooden box with a foam inset[55]. Following instruction, trainees demonstrated improvements in speed of closure, quality of closure and safety of closure using this model. A Medtronic (Medtronic Sofamor Danek, Inc., Memphis, TN, USA) ACDF model was developed by Ray et al to simulate this operation[52]. With pre-test, instruction and post-test, this model improved trainee performance using a custom assessment tool[52]. The two MIS models varied markedly in their design but both improved trainee performance and confidence[53,57]. The first model described by Walker et al developed a hybrid model using a deer lumbar spine that was affixed to components of a Halo orthosis frame and placed within a Plexiglas box to simulate MIS. Although only self-assessment was performed as part of this study, there was improved confidence in MIS skills[53]. The second model described by Chitale et al used a custom spine model with 3D navigation combined with pre-test, instruction and post-test[57]. Fluoroscopy time and number of shots improved, while there was a trend towards improvement in screw placement[57].

Cadaver simulation model

There were six studies which utilised cadaver models for the training simulation, with four studies assessing screw placement[40,42,58,59], one study each for dural closure[60] and spinal decompression with fusion[61]. Of the studies assessing screw placement, they ranged from cervical lateral mass screws[42] to thoracic pedicle screws[59] to lumbar pedicle screws with both Sundar et al and Podolsky et al examining all three regions within one educational session[40,58]. The educational session involved a three-day course covering cervical, thoracic and lumbar spine surgery. Three of these four studies demonstrated significant improvement in screw placement and accuracy following the education session, with only Podolsky et al finding no significant differences when compared to the control group. Two studies[40,59] were able to demonstrate how the education session was able to provide longitudinal benefits to trainees. This was unique amongst studies as most protocols included only same-day assessment for all participating trainees, allowing for assessment of construct validity in proving model utility. Buchanan et al demonstrated that a custom cadaver model designed to mimic CSF leak in a cadaveric specimen can be effectively used for education[60]. They assessed a group of eight trainees who improved in both objective (time taken for closure) and subjective (confidence in completing) measures of a minimally invasive dural closure. Calio et al presented a unique approach in that they made use of the cadaver to perform multiple different surgical procedures, maximising the education potential of this valuable resource[61]. Through use of both an anterior and posterior approach, they were able to provide 17 trainees of varying skill levels with appropriate instruction and assessment on these skills. Following completion of the course, trainees had increasing confidence in performing the appropriate steps of the procedure.

Virtual reality simulation

Virtual reality simulation provides an immersive surgical training experience that can be individually tailored to the trainee's needs. Four of the included studies in this review utilised virtual reality simulation models, two of which focused on screw placement techniques[62,63], one lumbar laminectomy[64] and one lateral lumbar access to the spine[65]. The two similar studies published in 2018 by Hou et al [62,63] compared a virtual surgery training system (VSTS) directly to traditional training methods (procedure explanation and video instruction) for cervical and thoracic pedicle screw placement respectively. The VSTS involved using a hand-held haptic feedback device to place pedicle screws into a 3D

generated graphic spine (Figure 1.3).

Figure 1.3: The virtual surgery training system[62]

Both groups were then tested on screw placement by inserting real screws into a cadaveric spine with post instrumentation imaging subsequently assessed by 3 independent observers. The conclusion of both studies found that the VSTS group had significant improvement in accuracy with decreased pedicle breach in both cervical and thoracic screw placement. Chen et al similarly utilised a 3D simulation training program looking at lumbar spinal stenosis and posterior decompression techniques[64]. This study showed an improvement in knowledge acquisition for anatomy and pathophysiology of spinal stenosis with VR technology, with greater improvements noted in more junior residents. Finally, Luca et al used a 3D headset and haptic feedback devices to perform lateral access to the lumbar

spine[65]. They found that users who were inexperienced with the technique showed significant improvement between first and second attempts, with a decreased number of major errors observed. The authors also included 3 senior surgeons into a separate group and commented that the users age or previous experience (or lack thereof) with VR technology did not affect the learning experience and usability of the device[65].

Augmented reality

While VR replaces the physical world with an entirely virtual world, AR technology aims to change the physical world by overlaying information onto it. Luciano et al used the Immersive Touch (ImmersiveTouch, Inc) augmented reality system which combines real-time haptic feedback with high-resolution stereoscopic display[66,67] (Figure 1.4).

Figure 1.4: Immersive Touch[66]

Participants used a haptic stylus as a drill to place thoracic pedicle screws. The Immersive Touch system allowed them to have constant fluoroscopy while drilling and inserting screws

and provides realistic drill feedback. In their first study, trainees were given a training module where constant fluoroscopic guidance was used. These adjuncts were then removed in the testing module and therefore "learning retention" was assessed. Fifty-one fellows and residents were recruited and placed 301 pedicle screws in both practice and test sessions. The failure rate of screw placement reduced from 16.9% to 12.5% demonstrating learning retention[66]. In their second study, Luciano et al used the Immersive Touch system to assess percutaneous pedicle screw insertion[67]. Sixty-three fellows and residents were recruited and placed 126 needles. They again had a training session where they used the fluoroscopic guidance provided by the system. However, during the test session they were unable to use fluoroscopy, instead relying on the 3D anatomy and tactile feedback they had learnt. Subsequently, the number of needle placement failures was the same in both sessions[67].

Cost of educational tools

The majority of the educational tools that were included did not specify the cost incurred in manufacturing and developing the tool. Of those that provided costings, the most expensive was the funding for Gottschalk's comparison of cadaveric and sawbone lateral mass screw insertion, which was $12,000[42].

Assessment of trainees within each study

Of the 30 included studies, 19 of them used the method of an assessment tool to evaluate the performance of their educational tool. The most commonly used assessment tool was the Objective Structured Assessment of Technical Skills (OSATS), used in 10 of 19 studies

where assessment was performed[19]. The OSATS is a 5-point global rating scale which assesses 6 domains: respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of operation, use of assistants and knowledge of the specific procedure. It has been widely validated as a tool for assessment of surgical trainees[37,68]. Six studies used custom assessments or questionnaires to evaluate their educational tool. Five studies assessed pedicle screw placement based on accuracy of placement, and rate of breach. The Physician Performance Diagnostic Inventory Scale (PPDIS) was used in 4 studies where trainees ranked themselves as unsatisfactory, early learner, competent, proficient or expert. Likert scales were used in 2 studies[60,61]. Buchanan et al used a 5-point confidence Likert scale[60]. Using a cadaveric model to simulate CSF leak, neurosurgery residents used the 5-point Likert scale from 1 = no confidence, to 5 = high confidence to rate their confidence with durotomy repair pre and post using the model. The mean pre model score was 1.88 which improved to 4.25 after the simulation model. Calio et al used a 10 point Likert scale to self-assess confidence, satisfaction in training and perception of safety in spine surgery[61]. They again used a cadaveric model for performance of ACDF, posterior cervical spine and posterior thoracolumbar spine surgery. Residents then used a 10-point Likert scale from 1 = strongly disagree to 10 = strongly agree, to rate their confidence in the surgery, their satisfaction in training and the perception of safety in performing the surgery[61].

Timeframe of assessment

Of the 25 educational tools assessed in this review, the majority evaluated trainee performance on the same day as the intervention. This immediate assessment provides

timely feedback and allows for immediate reinforcement of learning. However, it may not fully capture the long-term retention of knowledge and skills.

One study incorporated a pre-test 48 hours prior to the intervention, enabling the assessment of baseline knowledge and skills. This pre-post design provides valuable insights into the impact of the educational intervention on trainee performance. Additionally, one study reassessed trainees one month after the intervention, offering a longer-term perspective on knowledge retention and skill acquisition. While this approach is beneficial for assessing long-term learning, it may not be feasible for all training programs due to time and resource constraints.

The remaining studies reassessed trainees between one and two weeks after the intervention. This timeframe allows for some consolidation of learning and skill development, while still providing a relatively immediate assessment of the impact of the educational intervention. However, it is important to acknowledge that longer-term follow-up assessments may be necessary to fully evaluate the durability of knowledge and skills.

DISCUSSION

Using a systematic search strategy this review identified the educational tools that are have been utilised for educating spine surgery trainees in neurosurgery and orthopaedics. This is the first systematic review of all spine education tools and techniques. Due to changes in work-hour regulations, increased focus on patient-safety and economical constraints have resulted in a growth in simulation as an alternative means of educating surgical trainees[69]. We found varying forms of simulation are used for spine education including cadavers, custom models, 3D printed models and virtual and augmented reality simulators (Table 1.1).

Comparison of models

The models used have varying advantages and disadvantages. 3D printing is a rapidly growing technology that has many applications in medical education[46]. The versatility to create a range of individualised anatomical structures holds immense potential in the development of simulation models. The main constraint of 3D printed models is the current limitation in printing materials, with difficulty replicating the natural biological structure of human tissue[46].

Cadaveric laboratories have long been shown to be useful in spine surgical training but can often have difficulties with limited availability, high cost and individual specimen variability[70]. There are also inherently different characteristics between cadaveric bone to living bone (i.e lack of bleeding). Furthermore, some specimens have anatomical factors such as osteoporosis or obesity that make them inappropriate for training as they are not standardised.

Virtual reality simulation provides an immersive surgical training experience that can be individually tailored to the trainee's needs. Advancements in computer programming have resulted in the development of highly realistic and more complex simulation models, which can incorporate interactive feedback by adding input devices (i.e haptic feedback devices) so the user can interact with the virtual world. Furthermore, VR education may have greater advantages to traditional teaching methods, illustrated by a study which compared VR technology with more traditional education methods (i.e mannequin and cadaveric) and showed improved efficiency of knowledge transfer and decreased cost for VR education[71]. This must be considered in the context of VR being expensive to obtain, develop and run,

and additionally, not necessarily available in all contexts, hence not allowing access for all trainees.

Custom designed simulators were the most common form of spinal educational tool. While the majority of these tools were focused on pedicle screw insertion [18,50-53], ACDF[52], lumbar microdiscectomy[49], cervical laminoforaminotomy[56] and dural repair[54,55] models have been made and tested ensuring a good coverage of different spinal surgery. Sawbone models were the next most common model and again pedicle screw placement was tested in the majority of these tools[17,39-43]. In one study sawbones were found to be equivalent to cadavers, which is important as sawbones are much cheaper than cadavers[51]. Cadaveric models, and the newer technologies of 3D printing, virtual reality and augmented reality have been less commonly used for the purposes of education for the reasons stated above. However, in our search we did identify a number of models that have been validated but not yet used on trainees as an educational model. This included a number of 3D printed models, computer simulation software, augmented/virtual/mixed reality simulators[72-76]. In time, these may provide trainees with excellent learning opportunities.

Providing education with didactic instruction followed by a form of simulation resulted in improved surgical technique. In 28 of the 30 included studies, the educational intervention resulted in improved surgical accuracy, shorter operative time and improved confidence (Table 1.1). The percutaneous thoracic pedicle screw insertion Immersive Touch augmented reality model did not allow trainees to use fluoroscopy during its test module[67]. The authors acknowledged that percutaneous spinal fixation “depends mainly on the ability to interpret

fluoroscopy images". Podolsky et al[58] use of a 3D computer navigation system as an educational tool did not demonstrate an improvement in thoracic or lumbar pedicle screw insertion despite a 3 day course with didactic sessions included. As one of the older studies included, potentially this was related to the simulator itself.

Procedure based assessment

While the various simulation tools described above offer valuable opportunities for trainees to enhance their spinal surgical skills, their implementation is not without associated costs. These costs can be significant and include both direct and indirect expenses. Direct costs encompass the initial investment in the development and manufacturing of surgical models, as well as ongoing maintenance and repair. Additionally, there may be costs associated with acquiring specialised equipment, such as virtual reality systems or robotic simulators. Indirect costs, although less tangible, can also be substantial. These include the time required for trainees to engage with the tools, which may divert them from other clinical or educational activities, often reducing the time they spend in the operating theatre. Furthermore, there may be costs associated with faculty time for supervision, feedback, and assessment.

When assessing trainees the vast majority of educational models used same-day assessment of trainees to evaluate their tool. This involves immediate feedback and assessment of the trainee's skills and knowledge, providing an opportunity for timely reinforcement and corrective guidance. The means of assessment was via OSATS or a custom assessment tool

based on the OSATS[19]. None of these had been rigorously developed or validated in an evidence-based manner. Procedure-specific rating scales offer a balanced approach between broad, overall assessments (global rating scales) and detailed, task-by-task checklists. They focus on skills directly relevant to a specific procedure.

To create these scales, a procedure is broken down into individual steps. Then, performance quality descriptions are developed for each step. Raters use a Likert scale to quantify performance at each step. This provides objective feedback on technical skills, unlike global ratings. Trainees receive specific feedback on strengths and weaknesses, pinpointing areas for improvement. After targeted practice, trainees can be reassessed using the same scales to track progress and identify new development areas. While these scales offer significant advantages, their development and use remain relatively limited.

While most surgical training programs have included competencies in their curricula, their assessment has been difficult. While some of the educational tools included in this review used a custom designed assessment tool based on the OSATS[56], no procedure-based assessment tool has been designed for any form of spine surgery. An objective assessment scale for spinal surgery will improve the means by which competencies are assessed and produce better trained spinal surgeons.

In order to be able to assess technical competency there needs to be an appropriate assessment tool. In general surgery, specifically laparoscopic general surgery, a number of

objective assessment tools have been developed[6,7,30,77,78]. These tools are feasible, reliable and validated for each of the operations they are designed to assess. When used in a standardised curriculum for a specific operation in a training setting, surgical trainees that used the objective assessment tool had a superior improvement in knowledge and technical skills than those that did not[6]. In spinal surgery, through this systematic review we have identified a significant gap in the educational landscape when it comes to an objective, procedure-based assessment tool.

Limitations of our study

We acknowledge that there are limitations in our study. There are methodological limitations when performing a systematic review. These include the exclusion of studies based on language, publication date and study design (reviews). We also acknowledge that there are likely other assessment tools that have been used in surgical education and research in spinal surgery that may not have been published in the literature. This can introduce bias in to the results, and also limit the generalisability of the outcome of a systematic review. However, our purpose was to identify if there was a procedure-based assessment tool that existed for lumbar microdiscectomy. In terms of that specific goal, we were unable to identify a paper that described such a tool, and thus we were able to proceed with development and validation of the lumbar microdiscectomy assessment scale (LMAS).

CONCLUSION

To the best of our knowledge this is the first systematic review of education in spinal surgery. Current educational tools include cadavers, 3D printed models, sawbone models, custom based simulators, augmented and virtual reality simulators. The majority of these have demonstrated improved trainee confidence, accuracy and speed with spinal surgical technique across all aspects of spine surgery. However, they either involve a financial or time cost, and were not widely available to all trainees and centres. Procedure-based assessment tools on the other hand have been shown to improve performance, can be available to all trainees and are inexpensive. No procedure-based assessment tools currently exist in spine surgery, and as competency-based training becomes the most common form of training, a need exists for a procedure-based assessment tool in spine surgery. In the next chapter we discuss our development of the Lumbar Microdiscectomy Assessment Scale.

CHAPTER 2: DEVELOPMENT OF AN OBJECTIVE ASSESSMENT SCALE TO EVALUATE PERFORMANCE DURING LUMBAR MICRODISCECTOMY: A DELPHI ANALYSIS

In the previous chapter we showed through systematic review that there are no procedure-based assessment tools currently utilised or developed in spinal surgery. In this chapter we develop a procedure-based assessment tool for evaluation of performance during lumbar microdiscectomy.

This chapter is in manuscript form with view to submission in a peer reviewed journal.

The findings described in this chapter were presented as an oral presentation at the Royal Australasian College of Surgeons Annual Scientific Meeting Surgical Education section in Christchurch on 7th May 2024. I was awarded the Jenepher Martin Prize for best presentation in the Surgical Education section of the meeting.

ABSTRACT

Purpose

Lumbar microdiscectomy (LM) is the most common spinal procedure performed worldwide. Despite surgical training transitioning towards competency-based training, there are no objective methods of assessing competency in performing LM. The aim of this study was to reach expert consensus to create an objective evaluation tool for assessment of LM.

Methodology

Essential steps for LM and descriptors for “poor”, “average” and “excellent” performance at each step was developed. 16 surgeon experts were consulted and a Delphi methodology was used to obtain consensus on how suitable the steps and descriptors of performance

were for including in the assessment tool. Responses were obtained until consensus was achieved for each of these descriptors.

Results

Twelve experts from multiple institutions participated in the study. After two rounds of evaluations a Cronbach a of 0.99 was achieved, indicating a very high level of consensus. Steps and descriptors of performance that 80% of experts rated as 4 were used to create the final objective evaluation tool.

Conclusion

Using a Delphi methodology expert consensus was achieved for the first time regarding the steps and descriptors of performance for LM. Subsequently an objective assessment tool for LM surgery was created. This tool can identify strengths and weakness in performance and allow competency-based assessment.

INTRODUCTION

Spine surgery is a rapidly expanding field in Western medicine[31]. Unfortunately, complications from these procedures can severely impact patients' quality of life. This has led to a surge in malpractice claims against spine surgeons, resulting in higher insurance premiums[32,33]. The case of Dr. Christopher Duntsch serves as a stark reminder of the potential consequences when poorly trained spine surgeons perform routine procedures[1]. Questions persist regarding Dr. Duntsch's training, limited experience, and supervision. While his case is undoubtedly an extreme outlier, even highly skilled surgeons performing complex spine surgeries may experience complications in up to 87% of cases[1,32]. Even for a relatively simple procedure like a lumbar microdiscectomy, complications are more likely to occur in public hospitals with trainee neurosurgeons compared to private settings with experienced consultants[35]. Therefore, it's crucial that neurosurgical and orthopaedic trainees receive adequate training to safely perform lumbar microdiscectomy surgery which is a very common spinal procedure. Lumbar microdiscectomy is an index spinal procedure which junior trainees are performing despite its significant risks. Thus, it is of even greater importance to ensure that training and assessment of performance of this operation is rigorous.

The apprenticeship model, originating with Halsted in the 19th century, has long been the standard for surgical training. In the past trainees have therefore learned and practised surgery simultaneously[7,31]. But in the 21st century, advancements in medicine and surgery,

shifting patient and societal expectations, revised work-hour regulations, and a move away from long-term experiential learning assessments toward competency-based evaluations are reshaping surgical training. Competency based training sets predefined standards, technical and non-technical, that trainees must achieve by completion of training[5]. While there is a growing movement towards competency-based assessment in surgical training, traditional assessment methods, such as objective structured clinical examinations (OSCEs) and viva examinations, continue to play a significant role. These methods, while valuable for assessing knowledge and communication skills, may not adequately capture the complex and multifaceted nature of surgical competence.

Furthermore, a significant reliance on case logbook numbers persists in surgical training. While case volume is undoubtedly important, it may not be the sole determinant of surgical proficiency[79,80]. The quality of surgical performance, as measured by factors such as technical skill, decision-making, and patient outcomes, should be given equal weight. These assessment tools provide only a snapshot of a trainee's performance and are therefore not necessarily an accurate reflection of competence.

The increasing complexity of spine surgery, coupled with the potential for severe complications, underscores the critical importance of rigorous training for neurosurgical and orthopaedic trainees. As spine surgery continues to evolve, it is imperative that future surgeons are equipped with the necessary skills and knowledge to perform these procedures safely and effectively. By providing comprehensive training programs that emphasize both theoretical understanding and practical experience, we can help to mitigate the risks

associated with spine surgery and improve patient outcomes. This includes exposure to a diverse range of cases, supervised operative experience, and ongoing mentorship from experienced surgeons. Competency-based training is establishing itself as the leading training model to improve trainee skill acquisition and therefore, competency-based assessment tools must be developed. Ideally, a competency-based assessment tool should be standardised, valid, and reproducible in a global context[5]. The earliest competency-based assessment tools were global assessment scores, the Objective Structured Assessment of Technical Skill (OSATS)[19] (Figure 2.1) and Global Operative Assessment of Laparoscopic Skills Global Rating Scale (GOALS)[20,23] (Figure 2.2). These tools have formed the basis for a number of more specific procedure-based assessment tools[21]. Procedure-based assessment tools have the advantage that they can be validated for the procedure by experts in that procedure, can be reproducible in a global context if developed rigorously and allow trainees to gain feedback on specific areas of the procedure that require improvement. These tools are feasible and validated for each of the operations they are designed to assess. When used in a standardised curriculum for a specific operation in a training setting surgical trainees that used the objective assessment tool had a superior improvement in knowledge and technical skills than those that did not[6].

	1	2	3	4	5
Respect for tissue	Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately with minimal damage
Time and motion	Many unnecessary moves		Efficient time/motion but some unnecessary moves		Economy of movement and maximum efficiency
Instrument handling	Repeatedly makes tentative or awkward moves with instruments		Competent use of instruments although occasionally appeared stiff or awkward		Fluid moves with instruments and no awkwardness
Knowledge of instruments	Frequently asked for the wrong instrument or used an inappropriate instrument		Knew the names of most instruments and used appropriate instrument or the task		Obviously familiar with the instruments required and their names
Use of assistants	Consistently placed assistants poorly or failed to use assistants		Good use of assistants most of the time		Strategically used assistant to the best advantage at all times
Flow of operation and Forward planning	Frequently stopped operating or needed to discuss next move		Demonstrated ability for forward planning with steady progression of operative procedure		Obviously planned course of operation with effortless flow from one move to the next
Knowledge of specific procedure	Deficient knowledge. Needed specific instruction at most operative steps		Knew all important aspects of the operation		Demonstrated familiarity with all aspects of the operation

Figure 2.1: OSATS Global Rating Scale[19]

	1	2	3	4	5
Depth perception	Constantly overshoots target, wide swings, slow to correct		Some overshooting or missing of target, but quick to correct		Accurately directs instruments in the correct plane to target
Bimanual dexterity	Uses only one hand, ignores nondominant hand, poor coordination between hands		Uses both hands, but does not optimize interaction between hands		Expertly uses both hands in a complimentary manner to provide optimal exposure
Efficiency	Uncertain, inefficient efforts; many tentative movements; constantly changing focus or persisting without progress		Slow, but planned movements are reasonably organized		Confident, efficient and safe conduct, maintains focus on task until it is better performed by way of an alternative approach
Tissue handling	Rough movements, tears tissue, injures adjacent structures, poor grasper control, grasper frequently slips		Handles tissues reasonably well, minor trauma to adjacent tissue (i.e., occasional unnecessary bleeding or slipping of the grasper)		Handles tissues well, applies appropriate traction, negligible injury to adjacent structures
Autonomy	Unable to complete entire task, even with verbal guidance		Able to complete task safely with moderate guidance		Able to complete task independently without prompting

Figure 2.2: GOALS Global Rating Scale[23]

We have performed a systematic review of the literature on educational tools in spinal surgery and to date a procedure specific assessment tool for lumbar microdiscectomy has not been developed. This would allow trainees the opportunity to have their performance

assessed objectively with constructive feedback that they can use to improve their execution of this surgery. By providing detailed feedback on specific areas of strength and weakness, this sort of tool would empower trainees to target their learning efforts and focus on improving their performance in areas that require further development. Identifying specific areas of weakness allows trainees to practice deliberately on these areas. Deliberate practice has been used in sports and music for decades. It is the concept of engaging in activities created specifically to improve performance in a specific domain.[81] A procedure-based surgical assessment tool would facilitate deliberate practice and improve performance of the operation. This individualised approach to training can significantly enhance the efficiency and effectiveness of the learning process. The aim of this study was to develop a procedure specific tool for lumbar microdiscectomy to be used by neurosurgical and orthopaedic trainees to gain competence in lumbar microdiscectomy surgery.

METHODS

Study Design

A group of surgeon experts were consulted and a Delphi methodology used to obtain consensus regarding the essential steps for lumbar microdiscectomy. Descriptors for “poor”, “average” and “excellent” performance at each step were also developed and consensus achieved for each of these descriptors. The Delphi methodology was chosen due to its ability to allow expert opinions to be obtained in an anonymous manner and reassessed until consensus is achieved. It is a low cost, reliable and anonymous method of achieving expert consensus. Because there is no physical meeting, no single expert exerts dominance over the group and therefore it truly is a consensus-based approach. Ethical approval for the study was obtained from the human research ethics committee – Liverpool Hospital, South West Sydney local health district (2021/ETH11525).

Selection of Expert Surgeons

Leading neurosurgeons in Australia, New Zealand and the United States working in tertiary referral or academic neurosurgery centres with proven educational experience were contacted. The surgeons chosen all had extensive experience in neurosurgical education and training and directly supervised neurosurgical trainees. This was an essential criterion as the tool we are developing is for use by neurosurgical trainees. All surgeons contacted were involved anonymously.

Delphi Methodology

A hierarchical task analysis was conducted with the aid of surgical textbooks, videos, and expert opinion from faculty neurosurgeons at Liverpool Hospital, Sydney, to develop a list of potential steps required for lumbar microdiscectomy. Descriptors of "poor", "average" and "excellent" performance at each of these steps and sub-steps were also defined. Potential errors during lumbar microdiscectomy surgery were included in description of "poor" performance. Using a Delphi methodology, 16 experts from multiple institutions were asked to rate on a Likert scale from 1-5 how suitable these steps and descriptors of performance were for inclusion into an objective evaluation tool for lumbar microdiscectomy surgery. Each participant was emailed a link to an online survey using Google Forms that asked to rate each step, sub-step and descriptor using a Likert scale from 1 to 5 as to how important they felt these components should be included in a final assessment tool. Reminders were sent via email after 4 and 8 weeks.

Consensus And Creation of The Assessment Tool

Determining consensus when using a Delphi survey is not standardised. Previous work in this area has used Cronbach's α as a measure of the internal consistency of responses and as a marker of the consensus among a group of experts[5,7,30]. A Cronbach α approaching 1.0 is associated with greater internal consistency and therefore consensus. A value greater than

0.8 is suitable for the purposes of an evaluation tool[30]. Therefore, we used this as the cut off for consensus in this study, and also required that more than 80% of the experts agreed or strongly agreed that the steps, sub-steps and descriptors should be included in the final assessment tool.

Data Analysis

Once all the responses from experts had been collated the mean and standard deviations for scores that were given to each step, sub- step and descriptors of performance were calculated. Using R Studio statistical software package (Boston, MA), the Cronbach's α was calculated to determine the internal consistency of responses of the expert panel and therefore consensus. For each step, sub-step and descriptor the percentage of responses that agreed or strongly agreed was then calculated. If greater than 80% of responses agreed or strongly agreed then that step, sub-step or descriptor was included in the final evaluation tool.

RESULTS

Of the 16 experts contacted to be included in the study 12 participated (75%). Nine of the expert participants were from Australia, from seven different hospitals, while there were two from New Zealand and one from the United States. After collection of responses from the first round of the survey, the Cronbach α for the entire survey was 0.99. For the steps and sub-steps, the Cronbach α was 0.93 and for the descriptors it was 0.98. Of the 25 steps and sub-steps at least 80% of participants either agreed or strongly agreed that 22 should be included into the final lumbar microdiscectomy assessment scale (LMAS). “Prone”, “Make a midline incision” and “Placement of retractor to expose lamina, spinous process and facet joint” had less than 80% of participants agreeing or strongly agreeing. The descriptors for the “Prone” sub- step also had less than 80% of participants agreed or strongly agree. Therefore, after reviewing the comments the new step “Prone position, ensuring pressure areas are attended to” as well as new descriptors, and “Skin incision and identify subperiosteal plane, dissecting appropriately” were made. The step “Placement of retractor to expose lamina, spinous process and facet joint” was changed to “Placement of retractor to expose lamina, spinous process” following comments from the experts.

These changed steps and sub-steps were sent as a survey to the 12 experts and after one further round, the same statistics were calculated. More than 80% of participants now

agreed or strongly agreed with the new steps and sub-steps as well as the descriptors. With these new steps, sub-steps and descriptors the Cronbach α was calculated. For the entire survey it was 0.99, for the steps, sub-steps it was 0.96 and for the descriptors 0.98. Therefore, with these changes the LMAS tool was made (Figure 2.3).

Candidate			Assessor			
Operative Step	**Performance**					**Score**
	1	**2**	**3**	**4**	**5**	
Patient positioning and preparation						
Prone, ensuring pressure areas are attended to	Unable to align patient with table and pads, unsafe to turn prone		Performed, however, some adjustments required		Performed swiftly, without prompting, no repositioning required	
Ensure preoperative level is confirmed using either appropriate surface landmarks or XR	Does not identify appropriate surface landmarks nor uses XR		Performed after delay; prompting required		Performed swiftly, without prompting	
Preparation and drape applied	Poor technique; Loss of sterility; Inadequate preparation / draping over planned incision sites		Patient's back is prepared and draped adequately with prompting		Patient's back is prepared and draped using good sterile technique	
Exposure						
Skin incision and identify subperiosteal plane, dissecting appropriately	Unable to perform, requires repeated guidance, multiple ineffective traumatic movements to tissue, excessive bleeding, dangerous dissection, damage to surround structures		Safe but uneconomical movements to expose the subperiosteal plane. Requires some guidance.		Performed skillfully, with efficient, deliberate and effective movements without guidance	
Correctly place retractor to expose lamina; spinous process	Unable to perform, ineffective placement of the retractor on tissue with poor visualisation of anatomy, causes trauma to surrounding tissues, requires repeated guidance		Appropriate placement of retractor with some guidance/repositioning		Performed skillfully, with efficient, deliberate and effective movements without guidance	
Appropriate level check with X Ray	Ineffective placement of clip to confirm level, wrong level identified		Placement of clip to confirm level with guidance		Efficient placement of clip to confirm correct level without guidance	
Laminotomy						
Using either drill and/or Kerrison rongeur perform laminotomy to expose ligamentum flavum	Unable to perform. Performed clumsily with poor technique. Repeated guidance required.		Performed adequately but not efficiently with some guidance required		Performed skilfully and efficiently without guidance	
Appropriate partial medial facetectomy; preservation of pars	Unable to perform. Performed clumsily with poor technique. Repeated guidance required		Performed adequately but not efficiently with some guidance required		Performed skilfully and efficiently without guidance	
Safe opening of ligamentum flavum and flavectomy to expose dura; nerve root	Unable to perform. Performed dangerously, with poor technique. Risk to dura/nerve root. Repeated guidance required		Performed adequately and safely but not efficiently. Awareness of risk to dura and nerve root. Some guidance required		Performed skilfully and efficiently without guidance	
Neural decompression; rhizolysis						
Lateral recess decompression; rhizolysis of nerve root	Unable to perform. Unable to identify anatomy of lateral recess, pedicle, nerve root. Poor/dangerous technique. Repeated guidance required		Identifies anatomy, decompression performed adequately but not efficiently with some guidance		Performed skilfully and efficiently without guidance	
Rhizolysis of nerve root from origin to foramen	Unable to perform. Unable to identify anatomy of lateral recess, pedicle, nerve root, foramen. Poor/dangerous technique. Repeated guidance required		Identifies anatomy, decompression performed adequately but not efficiently with some guidance		Performed skilfully and efficiently without guidance	
Discectomy						
Careful retraction of nerve root medially	Inadequate/dangerous retraction. Repeated guidance required.		Adequate and safe retraction with some guidance		Excellent exposure with safe retraction. Full respect for tissues. No guidance required	
Satisfactory management of epidural veins	Unable to control veins. Causes trauma to veins resulting in bleeding. Repeated guidance required		Able to control veins with some guidance. No trauma to veins/bleeding		Controls veins skilfully and efficiently without guidance	
Annulotomy with knife or dissector, retraction of nerve root to enter disc space	Unable to perform. Dangerous /poor technique. Repeated guidance required		Performed carefully and adequately with some guidance.		Performed skilfully and efficiently without guidance. Full respect for tissues.	
Retrieval of free fragments while retracting nerve root	Unable to perform. Dangerous /poor technique. Repeated guidance required		Performed carefully and adequately with some guidance.		Performed skilfully and efficiently without guidance.	
Inspection; haemostasis						

Decompression of nerve with free passage of dissector along nerve and into foramen	Unable to perform. Unable to identify appropriate discectomy end point. Dangerous /poor technique. Repeated guidance required		Performed safely and adequately but requiring further discectomy in order to adequately decompress nerve. Some guidance required.		Performed skilfully and efficiently without guidance and need for further discectomy	
Obtain satisfactory haemostasis	Unable to control bleeding. Places surrounding structures at risk in attempt to control bleeding. Repeated guidance required		Able to achieve haemostasis but not efficiently. Some guidance required.		Performed skilfully and efficiently without guidance	
Closure						
Fascial closure	Unable to perform. Inadequate closure. Repeated guidance required		Performed carefully and adequately with some guidance.		Performed skilfully and efficiently without guidance	
Total Score						

Figure 2.3: The Lumbar Microdiscectomy Assessment Scale

DISCUSSION

Procedure-based assessment tools, such as this tool designed for lumbar microdiscectomy, are important and useful for transitioning surgical training to a competency-based system. This shift towards competency-based education emphasises the acquisition of specific skills and knowledge, rather than simply completing a predetermined number of procedures.

In this study, we have developed an objective assessment tool for lumbar microdiscectomy using a rigorous Delphi methodology. By engaging a panel of expert neurosurgeons, from academic centres in three countries, through multiple rounds of anonymous surveys and feedback, we have achieved a high level of consensus on the key steps, sub-steps, and performance criteria that define a successful lumbar microdiscectomy. This tool provides a standardised framework for assessing trainee performance, identifying areas for improvement, which may improve the quality of surgical training.

In designing this tool, we observed very high level of internal reliability across the entire rating scale, steps and sub-steps, and descriptors, with Cronbach α values >0.9. This surpassed our threshold of 0.8, and exceeds the Cronbach α values in previous studies [5,7,30,82]. We have implemented rigorous quality control measures to ensure the validity and reliability of our assessment tool. A key aspect of this process involved obtaining a high level of consensus among the expert panel. We confirmed at least 80% of the experts who participated in the Delphi methodology either agreed or strongly agreed with every component of the assessment tool. This substantial level of agreement underscores the

robust nature of the tool and its potential to provide a reliable and objective measure of trainee performance in lumbar microdiscectomy. In the first iteration of the Delphi survey three steps and three descriptors had less than 80% agree or strongly agree. “Prone”, “Make a midline incision” and “Placement of retractor to expose lamina, spinous process and facet joint” each had 66.67% of participants agree or strongly agree that they should be included as steps in the assessment tool. The comments for each step were reviewed and modifications were made.

Prone positioning was regarded as a fundamental component of the procedure, and thus the Delphi panel experts assumed that trainee neurosurgeons and orthopaedic surgeons would already possess this knowledge. Therefore, this step was changed and merged with the “Ensure pressure areas are attended to” step to form the step "Prone position, ensuring pressure areas are attended to". The previous descriptors “Not performed” for poor performance, and “Performed after delay, prompting required, significant repositioning required” for average performance had 58.33% and 66.67% of respondents agree or strongly agree they should be included. Therefore, these were changed to “Unable to align patient with table and pads, unsafe to turn prone” for poor performance and “Performed, however, some adjustments required” for average performance. “Make a midline incision” was also felt to be assumed knowledge and hence this was merged with “Dissection to expose subperiosteal plane” to form the step “Skin incision and identify subperiosteal plane, dissecting appropriately”. The step “Placement of retractor to expose lamina, spinous process and facet joint” was changed after comments to “Placement of retractor to expose lamina, spinous process” as most experts felt exposure of facet joint was an unnecessary

step in the operation and can in fact be deleterious for novice surgeons. These changes were then put in to another survey and sent to the 12 experts. On this occasion the sub-steps and descriptors had more than 80% agree or strongly agree they should be included in the rating scale. The Cronbach α was once again >0.9 for all components of the rating scale indicating high internal reliability.

The Delphi process is a structured communication technique used to achieve consensus among a group of experts[83]. The process involves anonymous response from these experts to a series of questionnaires over several rounds. This anonymity is a major advantage, as it reduces social pressure and bias, encouraging more honest feedback. The method is also geographically flexible, allowing experts from different locations to participate without the need for in-person meetings. Its structured, iterative nature and controlled feedback loop help refine opinions and lead to a more informed consensus[84]. However, the process is not without its drawbacks. It is often time-consuming, making it unsuitable for urgent decisions, and there's a risk of participant dropout over the multiple rounds[85]. The success of the process heavily relies on a skilled facilitator to avoid introducing bias, and the lack of face-to-face interaction can stifle spontaneous debate. Furthermore, while designed to prevent a group decision, the process can sometimes lead to participants conforming to the majority opinion[85,86].

In our study using a Delphi methodology allowed anonymous expert opinion and made two rounds of surveys seamless. The ability to have expert opinion from around the world without meeting is a distinct advantage of Delphi[87]. Having experts provide their opinion in an anonymous fashion also allows them all to have equal input without having one more dominant expert exert more influence than the others[87]. This also reduces any potential bias

in the creation of the tool. It is a systematic approach that strengthens the validitiy of our tool. Seventy-five per cent of the experts contacted participated in the creation of our tool, which is higher than the 60-70% that has been typically seen in other health related Delphi analyses[5,88]. This also reduces bias and enhances validity of our tool[88].

Despite the systematic approach of the Delphi analysis and the high level of internal validity and consensus amongst our experts there are some limitations of our study and the Delphi methodology. A primary concern is the potential for investigator bias in shaping the scale of the questions. As the Delphi process progresses, subsequent rounds may inadvertently narrow the scope of inquiry, leading to the omission of critical steps, sub-steps, or detailed descriptors. To mitigate this risk, we employed a multi-faceted approach:

1. **Comprehensive Literature Review:** We extensively reviewed established neurosurgical textbooks to ensure a broad and comprehensive understanding of the surgical procedure.
2. **Expert Consultation:** We engaged in in-depth discussions with local neurosurgeons to gather additional insights and perspectives.
3. **Iterative Refinement:** The iterative nature of the Delphi process may inadvertently narrow the scope of inquiry, leading to the omission of critical steps, sub-steps, or detailed descriptors. To offset this, we initiated the Delphi process with a comprehensive list of over 30 steps, progressively refining it to 25 and ultimately 22 steps through multiple rounds of expert feedback.

Another potential limitation stems from the inherent selection bias associated with expert panel recruitment. While we strived to involve a diverse range of experts from renowned academic institutions across Australasia and the United States, it is possible that certain biases, such as geographic, institutional, or disciplinary biases, may have inadvertently influenced the results. For example, experts from specific regions or institutions may have had a stronger emphasis on certain aspects of the surgical procedure, potentially leading to a skewed perspective. However, the high participation rate among invited experts and input from surgeons from multiple institutions with different practices helps to mitigate this issue, suggesting a broad consensus among the selected panel.

CONCLUSION

We were able to successfully utilise the Delphi methodology to obtain multi-institutional expert consensus regarding the steps, sub-steps and descriptors of performance during a lumbar microdiscectomy. Using this, we then created the Lumbar Microdiscectomy Assessment Scale (LMAS)the first objective procedure-based assessment tool for this index operation in spine surgery. This is a key step in developing a truly competency-based surgical training program. LMAS has the potential to enable trainees to develop a greater understanding of the steps and performance standards required to be proficient in lumbar microdiscectomy. Additionally, it can be used to enhance training and shorten learning curves by allowing trainees to understand what parts of the procedure require improvement. This is a key component in Deliberate Practice. LMAS can also be used for credentialling unsupervised practice by comparing trainee performance scores experts.

However, prior to using LMAS for these roles, its reliability and validity needs to be tested.

This is the focus of the next chapter in this thesis.

CHAPTER 3: DETERMINING THE RELIABILITY AND VALIDITY OF AN OBJECTIVE ASSESSMENT SCALE IN EVALUATING PERFORMANCE DURING LUMBAR MICRODISCECTOMY

In the previous chapter we developed the Lumbar Microdiscectomy Assessment Scale (LMAS) a procedure-based tool to evaluate performance of lumbar microdiscectomy. In this chapter we use video analysis of novice and expert surgeons performing lumbar microdiscectomy to validate and assess reliability of the LMAS.

This chapter is in manuscript form with view to submission in a peer reviewed journal.

ABSTRACT

Purpose

We previously developed the Lumbar Microdiscectomy Assessment Scale (LMAS) using a Delphi methodology. We sought to assess the reliability and validity of this tool.

Methodology

Using two groups of surgeons, those who have performed >100 lumbar microdiscectomy, and those who had performed <10, we recorded video of their operations. We obtained 8 videos from each group and then had two fellowship trained spinal surgeons assess the videos using the LMAS and OSATS.

Results

The overall ICC for the LMAS tool was 0.99 (95% CI 0.97-0.99; p<0.0001) with a correlation coefficient between the two independent blinded surgeons of 0.95. A coefficient of 0.95 (p<0.001) was observed between rater 1 and rater 2, and a coefficient of 0.82 (p=0.0001) between the combined LMAS and OSATS score for both raters.

Conclusion

We demonstrated that there is excellent inter-rater reliability and construct validity for the LMAS. We also confirmed concurrent validity when comparing the LMAS and the OSATS.

INTRODUCTION

Lumbar microdiscectomy remains one of the most commonly performed spinal surgeries worldwide, representing a cornerstone of spine surgery[31]. This procedure is frequently employed to alleviate pain and neurological symptoms associated with herniated lumbar discs, a condition that can significantly impact quality of life. However, complications arising from spinal surgery can have severe and potentially life-altering consequences, including lower limb weakness, paralysis, and dysfunction of the bladder, bowel, or sexual function[34]. Given the potential for serious complications, it is imperative that spine surgeons undergo rigorous training to ensure the highest standards of patient care. By acquiring the necessary surgical skills, knowledge and judgment, surgeons can minimise the risk of complications and optimise patient outcomes. This includes developing a deep understanding of spinal anatomy, surgical techniques, and potential pitfalls, as well as honing their technical skills through extensive training and practice. Rigorous training programs should emphasise a patient-centred approach, prioritising patient safety and well-being throughout the surgical process.

While experienced surgeons strive to minimise the risk of complications through careful preoperative planning, meticulous surgical technique, and vigilant postoperative care, it is important to recognise that complications can still occur, even in the hands of skilled surgeons. Despite the many advances in surgical techniques and the expertise of well trained, experienced surgeons, complications can still occur in up to 87% of lumbar microdiscectomy cases. [1,34]. These complication rates can be even higher in public hospital

settings where patients are often operated on by neurosurgical or orthopaedic trainees[35]. Trainees, while under the supervision of experienced surgeons, may have less experience and expertise, which can increase the risk of complications. Additionally, public hospitals may have higher patient volumes and more complex cases, further contributing to the increased risk of complications[35,36].

Traditionally, surgical training has followed an apprenticeship model, whereby trainees gain experience through a combination of observation and active participation in surgical procedures[13]. As trainees progress through their training, they are gradually granted increasing levels of responsibility, starting with simple tasks and progressing to more complex surgical manoeuvres under the supervision of experienced surgeons[14]. This apprenticeship model has been a cornerstone of surgical education for centuries, providing trainees with hands-on experience and the opportunity to learn from experienced mentors[89]. However, this model has inherent limitations, as it relies heavily on the availability of surgical cases and the willingness of senior surgeons to dedicate time to training. This “see one, do one, teach one” approach has permeated medicine and surgery for over a hundred years[13,14,89]. The reliance on traditional apprenticeship models has led to inconsistencies in the way surgery is taught and performed, resulting in variability in surgical outcomes and patient experiences. To address these challenges and improve the quality of surgical care, a paradigm shift towards competency-based training has gained prominence[89].

Competency-based training focuses on the acquisition of specific skills and knowledge, rather than simply completing a predetermined number of procedures[78]. By defining clear learning objectives and performance standards, competency-based training ensures that

trainees achieve a consistent level of competence before progressing to more advanced procedures. This approach helps to reduce variability in surgical practice and improve patient safety.[89].

The Royal Australasian College of Surgeons (RACS) has moved towards competency-based modes of training, where surgical trainees move through the different training phases by completing the requisite competencies within each phase[90]. In neurosurgery, lumbar microdiscectomy is one such competency, however, the way in which this is assessed remains a sign off from a consultant, rather than an assessment of the component steps of the operation. Following individual performance of surgery, more often than not, trainees are not provided with formal feedback on areas of improvement or methods to improve performance.

Global assessment scales such as the OSATS[19] were the earliest form of competency assessment for surgical procedures. They allow assessment generically in any surgical procedure. They provide a reliable and valid form of assessment of surgical performance. However, these tools focus on global assessments, rather than detailed evaluations of individual surgical steps and techniques. As a result, trainees may receive vague feedback that does not pinpoint specific areas for improvement. Without identifying the precise steps or sub-steps where a trainee may be struggling, it becomes difficult for them to target their practice efforts effectively. This lack of specific feedback can hinder the development of essential surgical skills and may lead to ongoing performance deficits.

To address these limitations, procedure-specific assessment tools, such as the one we have developed for lumbar microdiscectomy, offer a more granular approach to evaluating surgical performance. By breaking down the procedure into its constituent steps and sub-steps, these tools enable a detailed assessment of each component of the operation. This level of specificity allows for targeted feedback, enabling trainees to identify their areas of weakness and focus their practice efforts accordingly. Some procedure specific assessment tools have been developed but they were initially institution specific[26,28,29]. This limits the utility of such a training tool with a national based training program such as in Neurosurgery and Orthopaedic surgery in Australia and New Zealand.

Palter et al. developed a procedure-specific tool for laparoscopic colorectal surgery that demonstrated both reliability and validity[5]. This tool has the advantage of being widely applicable, as it can be used to assess the performance of trainees across different institutions. Inspired by this successful model and others in the field of general surgery[7,30], we aimed to develop a similar procedure-specific tool for lumbar microdiscectomy surgery. This tool would provide a valuable resource for both trainees and educators, facilitating the acquisition of essential surgical skills and enhancing the overall quality of surgical training. We have designed a procedure specific assessment scale, the Lumbar Microdiscectomy Assessment Scale (LMAS) using a hierarchical task analysis and Delphi methodology. In doing so, we identified and gained multi-institutional consensus regarding the component steps and sub-steps as well as descriptors of performance for these specific to lumbar

microdiscectomy surgery. Within this study, we aimed to assess the validity and reliability of the LMAS to assess performance during actual lumbar microdiscectomy.

METHODS

Study design

A prospective single-blinded observational study design was used to evaluate the reliability and validity of the LMAS. Ethical approval for this study was obtained from the Human Research Ethics Committee – Liverpool Hospital of the South West Sydney Local Health District.

Participants

At three tertiary hospitals in Sydney, New South Wales, patients with sciatica from lumbar disc herniation scheduled to undergo a lumbar microdiscectomy were identified through daily review of the elective neurosurgery operating lists at these hospitals from January 2023 to July 2024. Patients scheduled to have lumbar microdiscectomy surgery were asked to provide informed consent to video record their operation. If the patient agreed then the primary surgeon performing the surgery was also consented to include the video recording of their operation in the study.

Surgeons were categorised based on their operative experience in lumbar microdiscectomy surgery. Experienced surgeons were defined as those who were post-fellowship and had

performed >100 lumbar microdiscectomies as primary operator to ensure they were well beyond the learning curve. Novice surgeons were those who had performed less than 10 lumbar microdiscectomy as primary operator. One surgeon could not provide more than two independent videos of surgery.

Assessment tools

We used the recently developed Lumbar Microdiscectomy Assessment Scale (LMAS) described in the previous chapter for this study. This is a procedure-specific rating scale to assess performance during lumbar microdiscectomy surgery. It provides a structured framework for evaluating various aspects of the surgical procedure, including patient positioning, dissection, discectomy, and closure. By using this rating scale, educators can objectively assess trainee performance, identify areas for improvement, and provide targeted feedback to enhance surgical skills. This tool was developed by obtaining consensus amongst a multi-institutional group of experts with regards to the important steps, sub-steps and descriptors of “poor”, “average” and “excellent” performance to include in an evaluation tool. The final LMAS consisted of the agreed upon list of steps, sub-steps and descriptors of “poor”, “average” and “excellent” performance for each of these, scored on a Likert scale between one and five. The OSATS global rating scale was also used. This Likert scale assesses trainees on performance of surgery across the domains of respect for tissue, time and motion, instrument handling, assistant use, operative flow and knowledge of the procedure[19].

Video analysis

Video recordings were obtained from the operating microscope after the surgical site had been adequately exposed and before any bone removal procedures were undertaken. Inexperienced surgeons were closely supervised by experienced surgeons throughout the procedure. Both groups of surgeons were encouraged to maintain their usual surgical practices, including any guidance or advice that experienced surgeons would normally provide to junior colleagues. This approach aimed to minimise the impact of the video recording process on the surgeons' natural workflow and decision-making.

Video recordings were reviewed by two trained independent post fellowship surgeons to assess operative performance during each case using the LMAS and OSATS. These raters were post-fellowship surgeons, with an interest in surgical education, who had performed more than 100 lumbar microdiscectomies. The raters were blinded to the experience level of the surgeon. Both raters were allowed to familiarise and calibrate themselves with each evaluation tool. During this time instructions were given on the use of the descriptors of performance and Likert scale for each step. Videos could be played at faster playback speeds if it did not interfere with accurate evaluation of performance. The raters were asked to mark any steps within the rating scales but not visualised on the videos as “not applicable”.

Data analysis

The minimum number of video recordings required for each group was determined based on previous study that evaluated the Global Operative Assessment for Laparoscopic Skills (GOALS). In this study, the minimum relevant difference in mean scores between novice and

experienced surgeons was 6.4. Based on a standard deviation of 4.5, a power of 80% and an alpha of 0.05, the minimum number of video recordings required in each group was 8. All statistical analyses were performed using R studio statistical software package (Boston, MA). Descriptive statistics were calculated for the LMAS scores using medians and interquartile ranges.

Inter-rater reliability and test-retest reliability were used to assess reliability of the LMAS. Inter-rater reliability was determined using the intraclass correlation coefficient (ICC, 2-way mixed-effects model, absolute agreement), a measure of the internal consistency of the total scores and scores for each step as rated by the two blinded surgeons. A cut-off value of an ICC>0.8 has been acknowledged to be a benchmark for demonstrating good reliability[26]. Spearman's correlation coefficient was also used to assess correlation between the two raters. Test-retest reliability was evaluated by correlating the total score for a single case rated at 2 different points in time (index assessment and 3 months later) by each rater using ICC.

Mann-Whitney U test before and after pooling the scores were analysed for the two raters for each step to detect differences in experience level between the groups and establish construct validity. Spearman's correlation coefficient was used to correlate the LMAS scores with the corresponding OSATS scores. This allowed for assessment of concurrent validity. For the analyses of validity, each rater's scores as well as the combined LMAS scores were expressed as a percentage of the total possible score. Similarly, a percentage of the total score was used for the combined OSATS scores. This was conducted in case of a failure to capture or an inability to analyse part of a procedure.

RESULTS

Sixteen video recordings of lumbar microdiscectomy operations were obtained comprising of eight by novice surgeons and eight by expert surgeons. There were seven different experienced surgeons and five different inexperienced surgeons. For all surgeons, the performance rating for the step “Patient positioning and preparation” and the sub-step “skin incision and identify subperiosteal plane, dissecting appropriately” were excluded as this was not visible on the video. The LMAS was calculated out of a score of 60 and the OSATS out of a score of 35.

Reliability assessment

The overall ICC for the LMAS tool was 0.99 (95% CI 0.97-0.99; $p<0.0001$) with a correlation coefficient between the two independent blinded surgeons of 0.95. Each individual step of the LMAS tool was also analysed and the ICC was calculated for each with a range of 0.84 to 0.96 (table 3.1).

Component of LMAS	N	ICC	95%CI	P value
Correctly place retractor to expose lamina; spinous process	16	0.92	0.79-0.97	<0.0001
Using either drill and/or Kerrison rongeur perform laminotomy to expose ligamentum flavum	16	0.91	0.76-0.97	<0.0001
Appropriate partial medial facetectomy; preservation of pars	16	0.84	0.62-0.94	<0.0001
Safe opening of ligamentum flavum and flavectomy to expose dura; nerve root	16	0.84	0.61-0.94	<0.0001
Lateral recess decompression; rhizolysis of nerve root	16	0.86	0.62-0.94	<0.0001
Rhizolysis of nerve root from origin to foramen	16	0.89	0.72-0.96	<0.0001
Careful retraction of nerve root medially	16	0.86	0.65-0.95	<0.0001
Satisfactory management of epidural veins	16	0.89	0.72-0.96	<0.0001
Annulotomy with knife or dissector, retraction of nerve root to enter disc space	16	0.89	0.72-0.96	<0.0001
Retrieval of free fragments while retracting nerve root	16	0.96	0.89-0.98	<0.0001
Decompression of nerve with free passage of dissector along nerve and into foramen	16	0.92	0.79-0.97	<0.0001
Obtain satisfactory haemostasis	16	0.92	0.81-0.97	<0.0001

Table 3.1: Intraclass correlation coefficient for each component of LMAS

The inter-rater agreement for the inexperienced surgeons was 0.90 (95% CI 0.61-0.98, p=0.0002), and for experienced surgeons was 0.89 (95% CI 0.57-0.97, p=0.0003). Test-retest ICC was calculated for the LMAS. For rater 1, this value was 0.91 (95% CI 0.78-0.98, p=0.0006), for rater 2, this value was 0.82 (95% CI 0.61-0.93, p=0.0006), when combined the two raters test-retest ICC was 0.92 (95% CI 0.76-0.99, p=0.003).

Validity assessment

When the LMAS scores of experienced and inexperienced surgeons was compared there were significant differences in median scores (52.5% for the combined LMAS score of inexperienced surgeons and 99.2% for the experienced surgeons with p<0.0001) (Figure 3.1 and table 3.2).

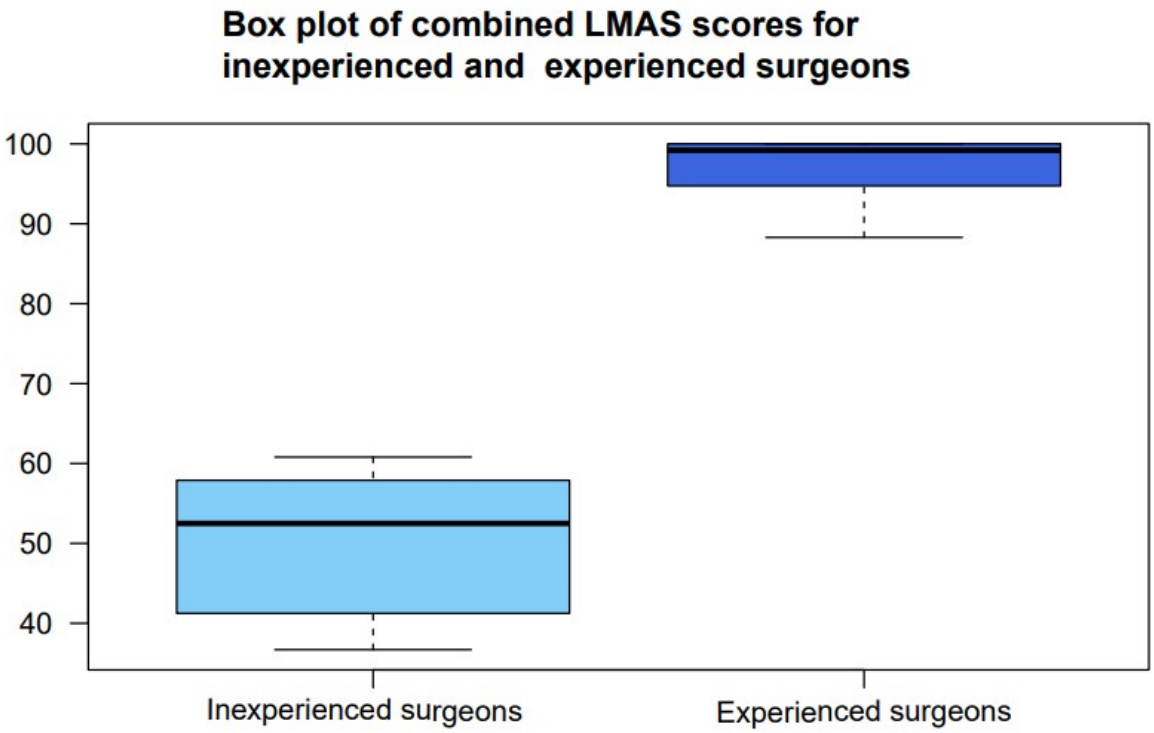


Figure 3.1: Box plot of combined LMAS scores for inexperienced and experienced surgeons

	Inexperienced surgeon	Experienced surgeon
Rater 1 LMAS	51.7% (33.3-60%)	100% (86.7-100%)
Rater 2 LMAS	53.3% (40-63.3%)	99.1% (90-100%)
Combined LMAS	52.5% (36.7-60.8%)	99.2% (88.3-100%)
Rater 1 OSATS	48.6% (37.1-68.6%)	94.3% (68.6-100%)
Rater 2 OSATS	47.1% (37.1-60%)	99.1% (85.7-100%)
Combined OSATS	46.4% (40-60%)	95.7% (82.9-100%)

Table 3.2: Scores on LMAS and OSATS for inexperienced and experienced surgeons

Spearman's correlation coefficient was calculated for concurrent validity. A coefficient of 0.95 ($p<0.001$) was observed between rater 1 and rater 2, and a coefficient of 0.82 ($p=0.0001$) between the combined LMAS and OSATS score for both raters (Figure 3.2 and 3.3).

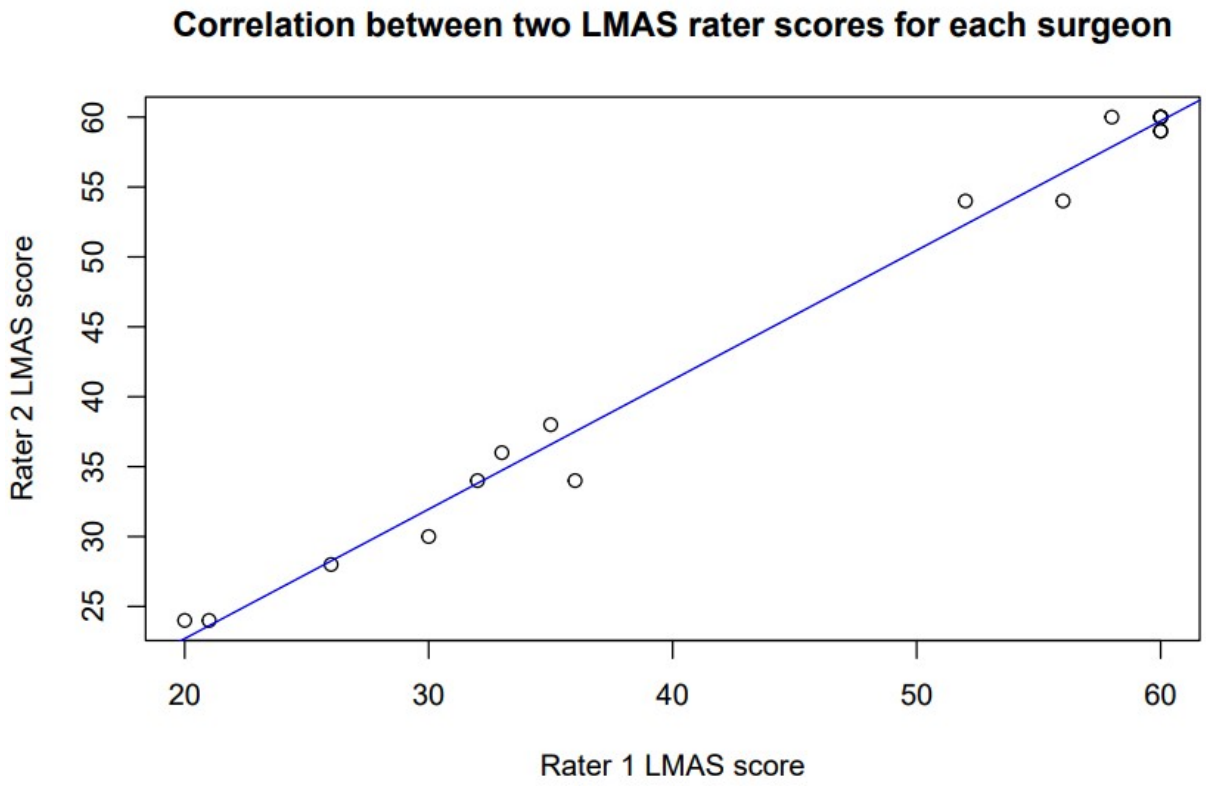


Figure 3.2: Correlation between the two LMAS raters for each surgeon

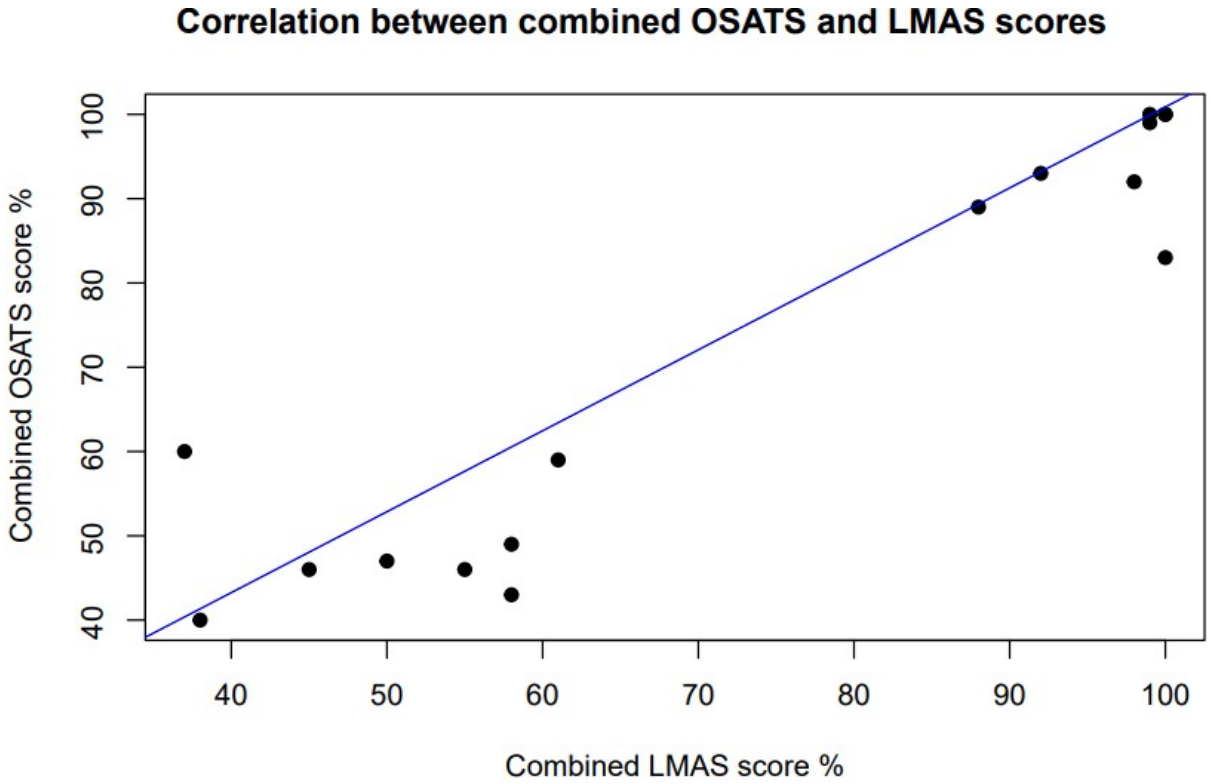


Figure 3.3: Correlation between the combined OSATS and LMAS scores

DISCUSSION

We have demonstrated that procedure-specific evaluation tool we developed for lumbar microdiscectomy surgery (LMAS) has excellent inter-rater reliability. We have also shown that this tool has construct validity on the basis that there were significant differences in the performance scores between novice and experienced surgeons. When compared with the gold standard OSATS tool, the LMAS scores correlated strongly indicating concurrent validity, confirming the LMAS tool can be used to accurately assess performance and most importantly, distinguish differences between varying skill levels. This procedure-based assessment tool could be used by trainees to evaluate their performance of lumbar microdiscectomy surgery and identify areas of weakness that could be worked on. We have established that the LMAS is reliable. The LMAS tool represents a reputable method of skills assessment to assist trainees in identifying their weakness and targeting their practice towards improving these. The tool has utility beyond trainees with potential to be used for

establishing competency benchmarks. These in turn can be used for performance goals or credentialing trainees for independent practice at lumbar microdiscectomy surgery, creating and fostering a competency-based training culture.

In our study we have demonstrated a higher reliability for the LMAS than that seen in studies of existing rating scales such as GOALS (ICC 0.89 for trained observers)[23], OSATS (Cronbach's alpha 0.72)[22] and modified OSATS (Cronbach's alpha 0.76)[22]. However, the reliability for the OSATS tool in our study was similar to that of the LMAS tool. Therefore, we believe the reason for our higher reliability is a function of the greater difference between the experience of the surgeons within the two groups. The inexperienced surgeons had performed less than 10 lumbar microdiscectomy surgeries, with novice skill levels, while the experienced surgeons had well over 100 lumbar microdiscectomy surgeries in their experience.

The LMAS has advantages over a global rating tool such as the OSATS tool as it delivers objective feedback on performance of specific steps relevant only to the performance of lumbar microdiscectomy. Therefore, this procedure-based assessment tool could be used by trainees to evaluate their performance during lumbar microdiscectomy surgery and identify specific areas of weakness that could be targeted for improvement. By reviewing the detailed feedback provided by the assessment tool, trainees can gain a deeper understanding of their strengths and weaknesses, enabling them to tailor their practice efforts and focus on developing the necessary skills to achieve optimal surgical outcomes. This tool empowers trainees to take an active role in their own learning and development

and gives trainers a framework with which to provide structured feedback and coaching, both of which foster a culture of continuous improvement in surgical education.

Global rating scales whilst easy to use and widely applicable are imprecise and therefore do not provide the level of feedback often required to improve technique for a specific procedure[25]. Procedure-specific rating scales confer a greater ability to differentiate between the performance of surgeons with different experience levels[91,92]. However, Birkmeyer et al demonstrated that a higher score on a global rating scale is associated with fewer postoperative complications, lower reoperation rates, and reduced readmission rates[24]. These findings suggest that a surgeon's technical skill and clinical judgment, as assessed by global rating scales, can significantly impact patient outcomes. Global rating scales allow overall surgical skill evaluation and so may have a role in overall certification at a board or fellowship level[8]. While procedure-specific rating scales, which focus on detailed assessments of surgical technique and performance, may be particularly well-suited for trainee credentialing and sign-off for independent practice in a specific operation, they may not capture the broader aspects of surgical competence[8]. These scales, while valuable for evaluating technical skills, may not adequately assess other important qualities, such as clinical judgment, communication skills, and problem-solving abilities[78].

To provide a comprehensive assessment of a surgeon's overall competence, it may be necessary to combine procedure-specific rating scales with more global assessments, such as those based on patient outcomes and patient satisfaction. By incorporating a variety of assessment methods, we can obtain a more holistic view of a surgeon's capabilities and identify areas for improvement and facilitate a comprehensive assessment [8].

All twelve of the components of the LMAS had an inter-rater reliability greater than 0.8 which is considered good or excellent[93]. This is testament to the consensus-based approach of the Delphi methodology and hierarchical task analysis that was performed to develop the rating scale. We believe the description of steps and sub-steps allows for ease of interpretation and our results support this. The calibration session held for the two raters also ensured clarity on what constituted each of the scores on the Likert scale from 1 to 5. This supervised joint calibration session ensured there was reliability when the blinded raters then rated the individual videos. By calibrating their ratings on a shared set of videos, the raters were able to establish a common understanding of the performance criteria and the nuances of the rating scale. This process helped to minimise variability in the ratings and enhance the overall reliability of the assessment. These “train the trainer” type sessions are important to the implementation of a procedure-based rating scale such as the LMAS.

There are some important limitations of this study. The use of video to assess performance retrospectively meant that the “Patient positioning and preparation” step and “Skin incision and identify subperiosteal plane, dissecting appropriately” sub-step of the “Exposure” step were not assessed. Similarly, the “Closure” step was also not assessed. This may have allowed the lower scores of the novice surgeons to increase. On a video it is difficult to know how much “guidance” is being provided verbally. While microscope video recordings capture only visual aspects of the surgical procedure, they do not fully capture the nuances of verbal communication, such as real-time feedback, explanations, and instructions. This limitation can potentially impact the objectivity of the assessment, as it may be difficult to determine

whether a particular surgical step was performed independently or with significant guidance from the supervising surgeon. A prospective intra-operative assessment would have helped circumvent these problems. However, this would unblind the raters introducing bias. The blinded nature of the raters to the surgeons is a significant strength of the study. By ensuring that the raters were unaware of the surgeon's level of experience, we were able to obtain objective and unbiased evaluations of surgical performance through video recording. There may have been instances where the consultant surgeon took over from a novice surgeon, and we acknowledge that this may have unblinded the assessment if the raters were able to ascertain a difference in glove colour or hand size between the two operators in the microscope view. However, this approach helped to mitigate the potential influence of factors such as surgeon reputation or institutional affiliation, which could otherwise bias the ratings. Despite this, both raters were still able to ascertain the difference between novice and expert surgeons' performance based solely on the video data available using the LMAS.

Requiring raters to watch a video for assessment may present a potential limitation of this tool. Even though videos can be watched on increased speeds, the sheer volume of footage can pose a significant time commitment. In our study, some videos, particularly those recorded by less experienced surgeons, exceeded an hour in length. This highlights the importance of efficient video review strategies. Zevin et al in their development of the BOSATS scoring system for Laparoscopic Gastric Bypass surgery were able to evaluate a 120-minute video in 30 to 40 minutes[7]. Time has also been raised as a limitation by others who have developed procedure-based assessment tools[26]. We interpret this as a valuable opportunity for trainees and surgeon educators to engage in constructive performance analysis. By utilising the LMAS, educators and trainees can objectively review video recordings of surgical procedures and identify specific areas for improvement away from the

immediate intraoperative pressures. This collaborative approach fosters a culture of continuous learning and enables trainees to develop the necessary skills to become proficient surgeons. The LMAS represents an invaluable learning opportunity and tool to allow objective assessment of performance, with the goal of improved surgeon training and ultimately patient care and outcomes[22].

There were seven expert surgeons and five novice surgeons that made the 16 videos available for analysis. Recruitment was challenging because of the very small pool of novice surgeons with the ability to perform lumbar microdiscectomy with supervision who had completed fewer than 10 procedures. This may have allowed some bias with the video recordings with potentially similar errors being noted by the raters for the novice surgeons that recorded two videos for inclusion in the study. The impact of this is minimal, as the LMAS scores for the performance of the novice surgeons remained consistent.

This study was conducted primarily at one institution, with two novice videos obtained from an external institution, challenging the generalisability of the study results. However, the LMAS was developed using the Delphi methodology with surgeons from Australia, New Zealand and the United States contributing to its design. This multinational input was intentionally included in the design process so the tool could be widely applicable. Our results support wider generalisability with two raters finding consistency in the ratings of novice and expert surgeons, but also a higher level of consistency with the LMAS compared to the OSATS for the same surgeons. For this reason, the LMAS has potential to be utilised in

surgical training and assessment widely. Further reliability and validity testing of the LMAS should be completed other institutions, in addition to correlation assessments.

CONCLUSION

We have demonstrated that a tool developed with expert input from around the world on a specific procedure, the lumbar microdiscectomy, can reliably evaluate performance of the operation in a reproducible manner. The rating scale was able to differentiate novice and expert surgeons. Therefore, this procedure-specific assessment tool has the potential to provide spinal surgical trainees with valuable feedback on their performance of the various steps and sub-steps involved in lumbar microdiscectomy surgery. By using the LMAS to evaluate their surgical videos, trainees can identify areas of strength and weakness and target their practice efforts accordingly.

Furthermore, this tool can be used by novice trainees who are eager to learn lumbar microdiscectomy. By reviewing the detailed performance criteria outlined in the LMAS,

novice trainees can gain a clear understanding of the specific skills and knowledge required to achieve competency in lumbar microdiscectomy surgery. This can help to guide their learning and provide a benchmark for assessing their progress. Specific feedback on their performance of the steps and sub-steps allows them to focus their attention, reading and practice on bettering those specific areas of weakness on assessment. A trainee's individual progress can therefore be tracked by using the LMAS at varying timepoints during their training. The LMAS can also be used as a final competency-based assessment to credential a trainee surgeon as experienced and independent in the operation. The LMAS could be utilised in the RACS Neurosurgery Surgical Education and Training program to ensure that trainees understand the steps and sub-steps involved in lumbar microdiscectomy within this new competency-based program. By promoting a culture of continuous learning and self-assessment, the LMAS can empower trainees to develop the necessary skills and knowledge to become proficient and confident spine surgeons.

CHAPTER 4: DISCUSSION AND CONCLUSIONS

In the previous chapter we validated the Lumbar Microdiscectomy Assessment Scale and demonstrated its reliability. In this chapter we discuss our findings and the implications of this thesis on competency-based education within spinal surgery.

INTRODUCTION

Spine surgery is an ever-evolving field of medicine which has impacts on up to 80-90% of the world[35]. Lumbar microdiscectomy is the most commonly performed spinal surgery worldwide[35,94]. The operation has been taught in the traditional "do one, see one, teach one" apprenticeship model of surgical training[31]. Traditionally, surgical training has relied heavily on unstructured, subjective methods of teaching, such as apprenticeship models. While these methods have historically been effective in transmitting surgical skills and knowledge from experienced surgeons to trainees, they lack the rigor and objectivity required for modern surgical education.

The aims of this study were to review the competency-based education tools available in spine surgery and identify if any procedure-based assessment tools existed. Identifying that there were no such tools, I aimed to develop one for lumbar microdiscectomy, which is the most commonly performed spinal procedure worldwide. A hierarchical task analysis and Delphi analysis were to be utilised to achieve this. Finally, once the tool was developed it a validation study was conducted.

In Chapter 1, I systematically reviewed the literature on educational tools and models that are utilised in spinal surgery training. These include saw bone models, 3D printed models, custom designed simulation models, cadaveric simulation models and virtual reality simulation models. Despite the complexity of some of these models, none of the currently available assessment and learning tools within spinal surgery are procedure based. Procedure-specific assessment tools offer a more detailed and unbiased way to evaluate surgical performance. By breaking down a surgical procedure into the constituent steps, these tools allow for a thorough assessment of each part of the operation. This specific feedback helps trainees pinpoint areas where they need improvement and focus their practice efforts.

Having identified that there are no procedure-based assessment tools currently available for use in training spinal surgeons, in Chapter 2 I developed the Lumbar Microdiscectomy Assessment Scale (LMAS) using a hierarchical task analysis and Delphi methodology. While procedure-specific assessment tools have been developed for other surgical specialties[5,7,26], the LMAS represents one of the first such tools specifically designed for spinal surgery. This innovative tool could provide a standardised framework for evaluating the performance of trainees in lumbar microdiscectomy, addressing a significant gap in surgical education and assessment. However, in order to confirm this, I needed to validate and assess the reliability of this tool using novice and experienced surgeons performing the operation.

In Chapter 3, I validated this tool using it to assess eight novice surgical videos and eight expert surgical videos in an independent blinded fashion. In doing so, we demonstrated that

the LMAS can be used in a validated fashion to assess trainee surgeons performing lumbar microdiscectomy surgery and identify strengths and weaknesses in their operative technique. This granular level of feedback can provide trainee surgeons with specific areas to practice and improve their performance. By identifying their weaknesses, trainees can target their practice efforts and focus on developing the necessary skills to achieve competency in lumbar microdiscectomy. This targeted approach to training can significantly accelerate skill acquisition and improve overall surgical performance. Furthermore, by setting clear performance benchmarks and providing regular feedback, procedure-specific assessment tools can motivate trainees to strive for excellence and achieve a high level of surgical proficiency. This can ultimately lead to improved patient outcomes and enhanced patient satisfaction.

IMPLICATIONS OF THIS RESEARCH FOR COMPETENCY-BASED MEDICAL EDUCATION

Several factors have contributed to the need for competency-based frameworks in surgical training. One significant factor is the reduction in training time for surgical trainees. As healthcare systems evolve and patient needs change, there is increasing pressure to shorten training programs while maintaining high standards of surgical competence. Competency-based training offers a solution to this challenge by focusing on the acquisition of specific skills and knowledge. By setting clear learning objectives and performance standards, competency-based training allows trainees to progress through the training program at their own pace, optimising their learning experience and maximising their potential. Work hour regulations, introduced first in the 1980s for patient and doctor safety now limit shifts to no

longer than 16 hours in Australia and New Zealand, with total hours limited to 72 hours in New Zealand since 1985[95]. While this has not significantly affected surgical volumes in training, it does mean that trainees have less time available in order to achieve the volume of cases they previously did. This is where competency-based training allowing trainees to achieve competency in a particular operation at their own time is a more suitable approach.

A longitudinal time-based approach to training has inherent limitations, primarily that it may not adequately account for individual differences in learning pace and skill acquisition[89]. Therefore, in recent years, there has been a shift towards competency-based training, which focuses on the acquisition of specific skills and knowledge. This approach emphasises the development of core competencies, such as technical skills, clinical judgment, and communication skills, rather than simply completing a predetermined number of years of training[89,96]. By setting clear learning objectives and performance standards, competency-based training allows trainees to progress through the curriculum at their own pace, optimising their learning experience and ensuring that they are adequately prepared for independent practice[8].

Competency-based training establishes clear, predefined standards, both technical and non-technical, that trainees must meet by the end of their training program and is a bridge between the didactic methods of the past and the current climate of medicine[15,30]. The core principles of competency-based medical education (CBME) are[97]:

- Establishing a competency framework

- Emphasis on assessment and feedback
- Individualised learning pathways
- Lifelong learning and continuing professional development

As previously discussed, surgical education has already established a competency framework and there is an emphasis on assessment and feedback. This is where the LMAS is best positioned within the principles of CBME. The LMAS is a tool that allows a consistent, valid and reliable assessment of the performance of lumbar microdiscectomy surgery. It identifies areas of weakness so that trainees can use Deliberate Practice to improve those specific areas. Thereby, individualising their own learning pathway. As discussed earlier, there are also opportunities with the LMAS to use it beyond trainees and training and as a credentialing tool for specialist spine surgeons to ensure there is a set reproducible standard at which the lumbar microdiscectomy is performed. In this way, it has a place within a lifelong continuous medical education framework.

Other such competency based tools are already in existence such as the OSATS[19] and GOALS[23]. These tools provide a reliable, feasible and valid assessment of surgical skills. The OSATS assesses six universal surgical skills assessed using a Likert scale:

1. Respect for tissue
2. Time and motion
3. Instrument handling
4. Knowledge of instruments
5. Use of assistants
6. Flow of operation and Forward planning

7. Knowledge of specific procedure

While these are very important “universal” surgical skills and widely applicable, they do not provide a trainee performing a lumbar microdiscectomy with specific feedback on if for example, the nerve root was adequately decompressed in the lateral recess. This is where a procedure-specific tool such as LMAS is more powerful for providing trainees with specific feedback on their performance of a specific operation, in this case lumbar microdiscectomy. They can identify that they scored poorly in “Safe opening of the ligamentum flavum” and tailor their practice and discussion with supervisors around this specific aspect of the operation. In doing so, they can improve that part of the procedure and progress towards achieving competence in performing the operation. In our analysis, the LMAS correlated strongly with the OSATS as described in Chapter 3, demonstrating concurrent validity.

ONTOLOGICAL AND THEORETICAL FOUNDATIONS OF THIS RESEARCH

This research is based on post-positivist pragmatism, which aims for objective assessment in surgical learning by balancing measurable results with the impact of contextual factors[9]. While acknowledging that complete objectivity is impossible, this approach seeks to approximate it through practical utility, reproducibility, and evidence-based methods. In the context of the LMAS, this meant using a structured, procedure-specific framework with predefined criteria to minimise judgment variability and ensure consistent feedback. This practical focus maintains rigour while acknowledging the complexities of surgical training.

Post-positivist pragmatism also addresses assessment fairness and reliability. Unlike traditional approaches that ignore context, this perspective incorporates contextual influences into tool design. For example, the LMAS uses standardised performance criteria to reduce subjective interpretation across different settings. By combining the pursuit of objective assessment with a flexible, outcomes-based framework, post-positivist pragmatism argues that reliable, actionable, and scientifically sound evaluations are achievable in real-world surgical training. This theoretical foundation is crucial to this thesis' goal of creating a robust and context-sensitive assessment for lumbar microdiscectomy.

POSTIONING LMAS WITHIN THE CONTEXT OF WORK-BASED ASSESSMENT IN SURGERY

The Royal Australasian College of Surgery in its practical guide to Work-based Assessment (WBA)[98] describes two forms of assessment – summative and formative. Summative assessment utilises a systematic, standardised evaluation process, characterised by the application of rigorous, evidence-based standards, to ensure the accurate and reliable assessment of participant performance and facilitate informed judgments regarding pass/fail outcomes[98]. While formative assessment strategies incorporate a set of rigorous standards as an integral component of a comprehensive continuing professional development or lifelong learning framework to facilitate the identification, through self-assessment or external evaluation, of areas where further development or skill enhancement is required[98]. WBA is defined as a “a central and integrating concept for the development of an assessment system whose purpose is both formative (educational) and summative (high stakes)”[98]. Miller’s work outlines this pyramid of assessment types and trainee performance

(figure 4.1)[99].

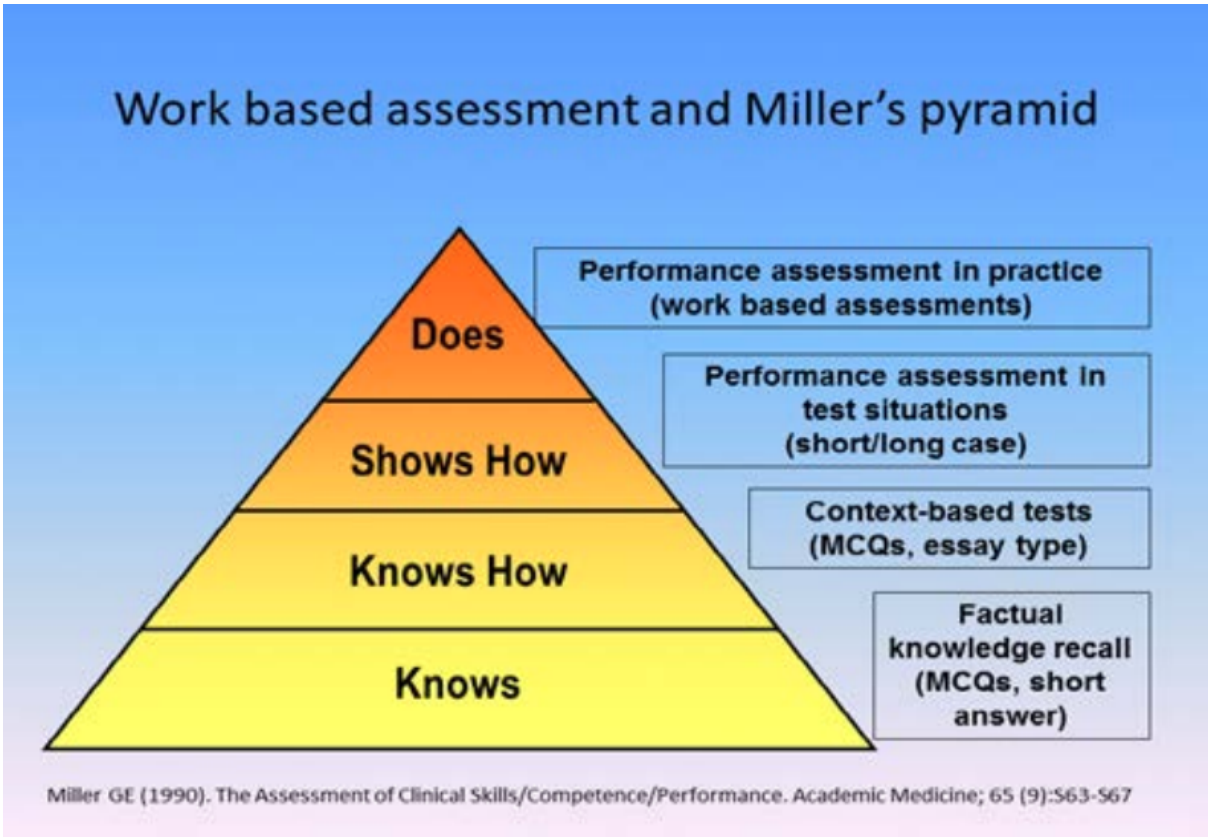


Figure 4.1: Work based assessment and Miller's Pyramid[99]

The three pillars of WBA[98] are that it

1. is a competency-based assessment, which is a requirement to assess trainees within a competency-based training curriculum
2. allows assessment against a set of external measures of performance, obviating self-assessment which in medicine is known to be inaccurate
3. is in line with public expectation of continuing professional development and medical education

WBA includes Procedure-Based assessment (PBA), Clinical Evaluation Exercise (CEX), Case-Based Discussion (CBD), Direct Observation of Procedural Skills (DOPS)[100].

The main criticism of WBA is that it does not often achieve the first two goals in a reliable fashion[101]. In a recent systematic review of WBA, Aryal et al found only 23 studies that evaluated WBA in surgery[100]. The vast majority of the studies assessed PBAs, with only 1 study on CEX, 2 on CBD and 3 on DOPS. Most of the included papers demonstrated that WBAs had good validity, reliability and acceptability. However, in a survey of 235 trainees and 239 trainers, (76% and 62% response rate), a number of criticisms of WBAs were identified[102]. Over 90% of trainers reported providing feedback to trainees, while only 73% felt they received feedback. WBAs were perceived as an assessment of trainee learning, rather than an assessment tool for learning. This undermines the formative component of the WBA and instead, the majority of trainees felt it was a summative assessment. There were also significant differences in the perception of time spent being assessed as well as time taken to receive feedback[102].

The LMAS with its procedure-specific component can be utilised in both a formative and summative fashion. Novice trainees learning lumbar microdiscectomy can identify gaps in their current skillset when performing the surgery and use deliberate practice to improve upon this. Intermediate and senior trainees can use the LMAS as a summative form of assessment to sign off on their competency in this operation. While self-assessment may be subjective, it does allow trainees an opportunity to assess their own performance. The LMAS allows trainees to review their own videos and assess their own performance. Surgical assessors should also be able to perform assessment while watching their trainees in real time. There is still a time-based component to completing the LMAS at the completion of an operative case, and there is still the requirement for a trainee and supervisor to discuss their

score on the LMAS and identify areas of strength and weakness. However, this should not be any different to existing WBAs used for this procedure.

Trainers often perceive work-based assessment (WBA) as providing a more significant opportunity for learning compared to the perspective of trainees themselves. This disparity highlights a potential mismatch in perceptions regarding the delivery and receipt of feedback within the WBA framework[102]. My hope is that with the LMAS and its procedure specific nature, the trainees can identify areas in which they are performing the operation well as well as areas that require further effort. They can then use the LMAS in an ongoing fashion until such time that they achieve competency across all the components of the operation. Despite this, there will still be a consideration among trainers that the on-the-job context of WBA, including the LMAS, facilitates authentic learning experiences, allowing trainees to apply theoretical knowledge in real-world situations and receive immediate, relevant feedback from experienced practitioners[100,103]. This can foster deeper understanding and skill development. Conversely, trainees may perceive WBA as primarily focused on assessment and meeting the requirements for certification and more of a box-ticking exercise, which overshadows any potential learning opportunity[101]. Hopefully with the LMAS and the ability to use it along their journey of learning lumbar microdiscectomy surgery, they will not experience anxiety related to performance evaluation that they have previously endured with other forms of WBA[100-102].

This mismatch in perceptions between trainers and trainees can have significant implications for the successful implementation of WBA. It underscores the importance of clear communication, shared understanding of learning objectives, and a strong focus on

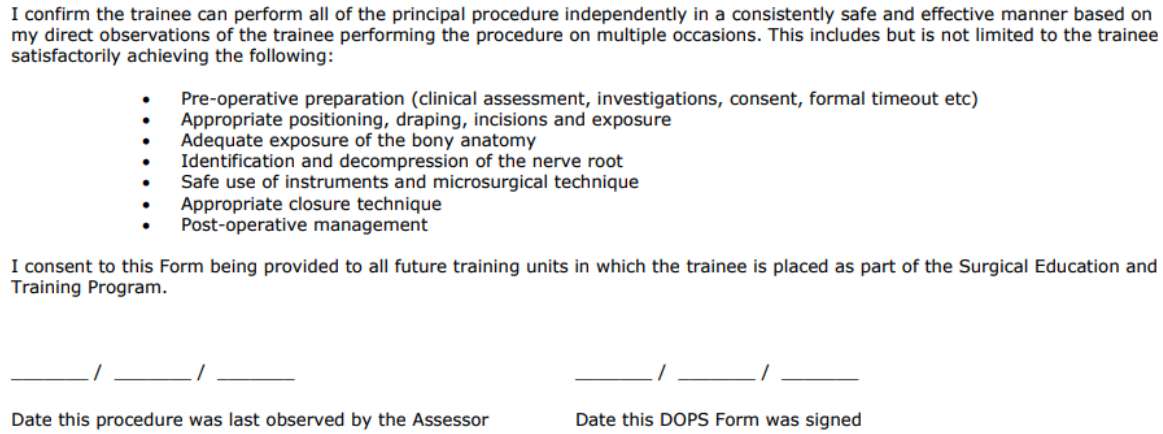
providing meaningful and actionable feedback to trainees throughout the WBA process[102,103]. Another significant concern expressed by trainees, particularly those with advanced levels of training and experience, was the perceived lack of specificity in the assessment criteria, which did not adequately differentiate performance expectations based on individual levels of expertise. Due to its procedure-specific nature, and the hierarchical task analysis breaking down the operation into its component steps, the LMAS directly addresses this concern. It is specific to the lumbar microdiscectomy procedure, with clear descriptors of performance. There was also high correlation between the two raters of the LMAS, which demonstrates that even with minimal training, the LMAS can be used in a reproducible fashion. This should obviate another common complaint with PBA, that that trainees feel trainers have not had adequate training in using PBA[101,102].

In the RACS General Surgery training program DOPS forms include generic feedback such as seen on this form (Figure 4.2).

Please assess and mark the following areas:	Below expectations for level of training	Borderline	Meets expectations	Above expectations for level of training	Not observed Or not applicable
1. Explains the procedure and complications to the patient and obtains patient's informed consent					
2. Prepares for procedure according to an agreed protocol					
3. Demonstrates aseptic techniques and safe use of instruments/sharps					
4. Performs technical aspects competently					
5. Demonstrates manual dexterity required to carry out procedure					
6. Adapts procedure to accommodate patient and/or unexpected events					
7. Is aware of own limitations and seeks help when appropriate					
8. Completes required documentation (written or dictated)					
9. Analyses one's own clinical performance for continuous improvement					

Figure 4.2: General Surgery DOPS

In the RACS Neurosurgery training program DOPS for Lumbar microdiscectomy the feedback is more specific, however, the way in which the components of the operation are rated is open to interpretation (Figure 4.3).



I confirm the trainee can perform all of the principal procedure independently in a consistently safe and effective manner based on my direct observations of the trainee performing the procedure on multiple occasions. This includes but is not limited to the trainee satisfactorily achieving the following:

- Pre-operative preparation (clinical assessment, investigations, consent, formal timeout etc)
- Appropriate positioning, draping, incisions and exposure
- Adequate exposure of the bony anatomy
- Identification and decompression of the nerve root
- Safe use of instruments and microsurgical technique
- Appropriate closure technique
- Post-operative management

I consent to this Form being provided to all future training units in which the trainee is placed as part of the Surgical Education and Training Program.

_____ / _____ / _____ _____ / _____ / _____

Date this procedure was last observed by the Assessor Date this DOPS Form was signed

Figure 4.3: Lumbar microdiscectomy DOPS SET Program in Neurosurgery

The LMAS provides clear descriptors of performance for all the component steps of the operation, making it clear for both trainee and supervisor the level of performance achieved by the trainee. This should provide trainees with an opportunity to target their areas of weakness until they achieve competence in the operation.

IMPACT ON PATIENT OUTCOMES AND SAFETY

The haunting example of Christopher Duntsch still lingers in spine surgery. Patients watching his documentary carry great fears regarding spine surgery and its outcomes, and it certainly adds to the negative attitude toward the profession that purveys the community. His was hopefully a singular example, however there were great concerns and questions raised regarding his training, and spine training in general[1]. More recently the man alleged to have

murdered a health care executive had a poor outcome from spine surgery also[104]. The overarching objective of our spinal surgical trainee education program is to cultivate a new generation of highly skilled and compassionate surgeons who are equipped to consistently deliver the highest quality of care to spinal patients, ultimately contributing to improved patient outcomes, reduced complications, and enhanced quality of life. The move towards competency-based training was aimed at achieving this.

Birkmeyer et al.'s landmark 2013 study provided compelling evidence that increased surgical skill significantly impacts patient outcomes[24]. Their findings revealed a strong correlation between higher levels of surgical expertise and reduced rates of postoperative complications, the need for reoperations, hospital readmissions, and emergency department visits following surgical procedures[24]. This was reflected in another more recent study of colorectal surgery[105]. Therefore, using a tool such as the LMAS which should improve trainee surgical skills during training can translate in to improved surgical outcomes.

This concept has been demonstrated in other specialties. An anaesthetic review on competency-based learning for insertion of central venous catheters demonstrated improved patient outcomes and resultant cost-saving from $63,000 up to $700,000 in one study[106]. In surgery there has been less evidence of this direct benefit. In their systematic review Meling and Meling found that operative time was reduced with competency-based training and on global rating scores, and simulation training improved scores[107]. Prolonged operating times are known to be associated with a number of complications including infection, thromboembolic phenomena, cardiorespiratory complications and renal

impairment[108]. Accordingly, the reduction of operative time, facilitated through the implementation of rigorous simulation training programs and a robust competency-based learning framework, can contribute significantly to enhanced patient outcomes and improved surgical safety.

Another example of a competency-based education program improving performance of surgery and patient outcomes is described in this urological study. In what is the first randomised controlled trials on competency-based education, Aydin et al randomised 94 urology trainees to simulation-based (SBT) and non-simulation-based training (NSBT) with a primary outcome of identifying the number of procedures to achieve proficiency in ureteroscopy, the index urology procedure[109]. The secondary outcome was the number of complications in each group. They found that overall proficiency in flexible ureteroscopy, a technically more challenging procedure, was achieved by 20 participants in the SBT group compared to only 9 in the NSBT group. Furthermore, there were fewer complications in the SBT group (15 compared to 37, p=0.003). This RCT furthers the concept that if a competency-based program or tool, such as the LMAS, can improve proficiency of a surgical skill, this translates in to fewer complications and therefore improved patient safety.

It has been demonstrated that enhancing the surgical skills of trainees can significantly impact surgical performance, ultimately leading to improved patient outcomes and enhanced surgical safety. The LMAS, with its demonstrated ability to differentiate between expert and novice surgeons, presents a valuable tool for surgical skill assessment and development. Trainees should be encouraged to utilise the LMAS throughout their surgical

training journey. By consistently tracking their LMAS scores and striving to achieve performance levels comparable to those of expert surgeons, trainees can actively engage in a process of continuous skill refinement and improvement. This iterative approach, facilitated by the LMAS, has the potential to significantly enhance the performance of lumbar microdiscectomy surgery, thereby minimising complications such as infection, nerve root injury, and prolonged recovery times. Ultimately, by fostering a culture of continuous skill improvement and utilising tools like the LMAS, we can contribute to improved patient safety and significantly enhance the overall quality of patient outcomes for this critical spinal surgical procedure.

STRENGTHS AND LIMITATIONS

There are a number of strengths in the development of the LMAS and its validation. In seeking expert opinion for creation of a rating tool such as the LMAS, there is potential for institutional bias, one expert to overrule others and potential for other biases to be introduced. Geographical differences in technique, can result in experts from these areas prioritising certain aspects of the procedure over others which can also lead to bias when developing a tool such as the LMAS. This would interfere in its ability to be widely applicable, reliable, valid and feasible as a tool to assess competency. To try and avoid this we used the Delphi methodology to develop the LMAS[87]. Using the Delphi method facilitated gathering anonymous expert opinions globally through two smooth survey rounds. This approach offers the significant benefit of collecting insights from international experts without requiring in-person meetings[87]. Anonymity ensures that all experts contribute equally, preventing any single dominant voice from unduly influencing the process and minimising

bias in the tool's development. Especially when there are institutional and even geographical preferences in terms of technical approaches to an operation, this Delphi methodology offsets this, and provided reliability from the outset in our development of the LMAS.

Another important component of the LMAS is the use of video recordings. This was a significant strength of the validation of the LMAS because no alteration to standard surgical management was required. Videos were recorded on the operating microscope used in all cases. While this was a clear strength of the LMAS validation, there were other considerations also. While the operating microscope is used for the majority of the operation, it is not introduced until after exposure of the spine and the level has been confirmed with x-ray. Therefore, the initial steps of “Patient positioning and preparation” and “Skin incision and identify subperiosteal plane, dissecting appropriately”, could not be assessed in the validation study. Similarly, “Closure” was not assessed as it was not on the video. If these steps had been included it may have increased the lower scores of the novice surgeons. To offset this an intraoperative assessment could have been performed by our raters, however, this would unblind the surgeons and introduce bias. Similarly, on video it is sometimes difficult to assess the degree of “guidance” being utilised. Video analysis does not demonstrate all of the nuances of verbal communication, real-time feedback, explanations, and instructions. This can impact objective evaluation, as it can be challenging to distinguish between a surgical step performed independently versus one completed with substantial guidance from a supervisor. Again, intraoperative assessment would have allowed for assessment of this but unblinded the assessors reducing objectivity of assessment. This methodology minimises the impact of factors like surgeon reputation or

institutional affiliation, which could otherwise skew the ratings. Nevertheless, even with this mitigation, both raters were still able to differentiate between novice and expert surgical performance using the LMAS based solely on the available video footage.

Finally, the sample size of our study to validate the LMAS was small. This can result in questions regarding the generalisability of our results which is a valid concern. Other similar studies have used similar sample sizes, and the next steps would be to consider using the LMAS in other institutions with larger numbers to confirm the generalisability and correlation that we found in our study.

IMPLEMENTATION AND FUTURE DIRECTIONS

The aim of this thesis was to develop a procedure-based assessment tool for lumbar microdiscectomy that was valid, reliable and feasible for use in training surgeons who perform lumbar microdiscectomy. We have demonstrated that the tool is valid and reliable within the context of our study In achieving both goals, our next goals are that the LMAS can be introduced into spinal surgical training programs to improve trainee performance of lumbar microdiscectomy and thereby improvement of patient outcomes. Ultimately, ensuring there are no repeats of what happened with the patients of Dr Christopher Duntsch.

The first step for achieving this is to utilise the LMAS in other institutions and ensure reliability and generalisability of our tool beyond our study. With multi-institutional

confirmation of its use, this will also introduce the LMAS to more supervisors and trainees. This is important for the next phase of implementation which is introduction into the training curriculum for both the Neurosurgery and Orthopaedic surgery training programs. If this can be achieved, the approach used in this thesis can be applied to other spinal surgical procedures to replicate the LMAS for these, ultimately developing a surgical curriculum with procedure-specific assessment tools across all procedures.

Once further multi-institutional validation has occurred, the LMAS could be used to demonstrate that it improves performance of lumbar microdiscectomy surgery. Trainees could use it throughout their training and information such as the learning curve of lumbar microdiscectomy surgery can also be clarified. Furthermore, the LMAS platform can be effectively leveraged for summative assessment and credentialing purposes. By analysing the performance of expert surgeons using the LMAS tool, robust performance benchmarks can be established. These benchmarks serve as aspirational goals for trainees, providing a clear and objective standard for achieving surgical competence and proficiency in this operation.

Upon consistently demonstrating performance levels that meet or exceed these established benchmarks, trainees can be deemed competent in lumbar microdiscectomy surgery and granted the necessary credentials for independent practice. This data-driven approach fosters a culture of continuous improvement, ensuring that graduating surgeons possess the necessary skills and expertise to deliver safe and effective patient care.

One method of ensuring ease of access and enhancing the practical utility of the LMAS would be through the development of a user-friendly, cloud-based platform. This platform would feature individualised trainee accounts, providing secure and convenient access to performance data. During surgical procedures, assessors could effortlessly utilise a mobile device to score trainee performance using the LARS tool. This real-time data would then seamlessly uploaded to the trainee's account, creating a dynamic and readily accessible record of their surgical progress. With further refinement of the technology areas of weakness in performance could be identified and targeted interventions could be recommended for improvement. This may include enrolment in specialised skills labs, access to simulation-based training modules, or mentorship from experienced surgeons to address specific areas of weakness.

CONCLUSION

In developing the LMAS and demonstrating its validity I have adhered to the core principles of competency-based medical education. Using a competency framework the hierarchical task analysis and Delphi methodology outlined the steps of the operation that if performed at an “excellent” level on a Likert scale result in achieving competence at performing the surgery. In validating the tool, I have demonstrated that it can be used in a reliable fashion by trainees and the LMAS has a strong correlation with the OSATS when assessing trainee performance. In doing this the focus of the LMAS has been on assessment of performance and allowing supervisor feedback to improve this performance. The LMAS allows individualised learning and provides trainees with specific areas for improvement and

deliberate practice to achieve this. Beyond these initial phases of development and validation, there are broader goals for the LMAS.

If it can be validated more broadly in other institutions and implemented by trainees in these institutions to improve performance, I believe that there can be paradigm shift within spinal surgical education starting with lumbar microdiscectomy the index procedure in this specialty. Some of this future direction of this thesis include demonstrating the improvement of trainee performance using the LMAS. Demonstrating benchmarks for performance of lumbar microdiscectomy as well as the learning curve of the operation are also important next steps. In providing more data for this index spinal surgical procedure my hope is that there can be more broad reform of how we educate spinal surgical trainees. By integrating this innovative tool into the existing training framework, we can establish a foundational step towards a broader, nationwide transformation of surgical education in Australia improving patient outcomes along the way.

REFERENCES

1. Chapman JR, Wang JC, Wiechert K. How Should We Deal With "Black Swan" Surgeons in Spine Surgery? *Global Spine J*. Jun 2019;9(4):365-367. doi:10.1177/2192568219847207
2. Aggarwal R, Grantcharov TP, Darzi A. Framework for systematic training and assessment of technical skills. *J Am Coll Surg*. Apr 2007;204(4):697-705. doi:10.1016/j.jamcollsurg.2007.01.016
3. Aggarwal R, Hance J, Darzi A. Surgical education and training in the new millennium. *Surg Endosc*. Oct 2004;18(10):1409-10. doi:10.1007/s00464-004-8238-x
4. Urbina J, Monks, S.M. Validating Assessment Tools in Simulation. *StatPearls*. StatPearls Publishing; 2025.
5. Palter VN, MacRae HM, Grantcharov TP. Development of an objective evaluation tool to assess technical skill in laparoscopic colorectal surgery: a Delphi methodology. *Am J Surg*. Feb 2011;201(2):251-9. doi:10.1016/j.amjsurg.2010.01.031
6. Shore EM, Grantcharov TP, Husslein H, et al. Validating a standardized laparoscopy curriculum for gynecology residents: a randomized controlled trial. *Am J Obstet Gynecol*. Aug 2016;215(2):204 e1-204 e11. doi:10.1016/j.ajog.2016.04.037
7. Zevin B, Bonrath EM, Aggarwal R, et al. Development, feasibility, validity, and reliability of a scale for objective assessment of operative performance in laparoscopic gastric bypass surgery. *J Am Coll Surg*. May 2013;216(5):955-965 e8; quiz 1029-31, 1033. doi:10.1016/j.jamcollsurg.2013.01.003
8. Aggarwal R. Intraoperative Surgical Performance Measurement and Outcomes: Choose Your Tools Carefully. *JAMA Surg*. Nov 1 2017;152(11):995-996. doi:10.1001/jamasurg.2017.0837
9. Varpio L, MacLeod A. Philosophy of Science Series: Harnessing the Multidisciplinary Edge Effect by Exploring Paradigms, Ontologies, Epistemologies, Axiologies, and Methodologies. *Academic Medicine*. 2020;95(5):686-689. doi:10.1097/acm.0000000000003142
10. Bunniss S, Kelly DR. Research paradigms in medical education research. *Medical Education*. 2010;44(4):358-366. doi:10.1111/j.1365-2923.2009.03611.x
11. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher*. 2009;29(9-10):855-871. doi:10.1080/01421590701775453
12. Rethans JJ, Norcini JJ, Barón-Maldonado M, et al. The relationship between competence and performance: implications for assessing practice performance. *Medical Education*. 2002;36(10):901-909. doi:10.1046/j.1365-2923.2002.01316.x
13. Skjold-Odegaard B, Soreide K. Competency-based Surgical Training and Entrusted Professional Activities - Perfect Match or a Procrustean Bed? *Ann Surg*. May 1 2021;273(5):e173-e175. doi:10.1097/SLA.0000000000004521
14. 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
15. Hanrahan JG, Sideris M, Pasha T, Dedeilia A, Papalois A, Papalois V. Postgraduate Assessment Approaches Across Surgical Specialties: A Systematic Review of the Published Evidence. *Acad Med*. Feb 1 2021;96(2):285-295. doi:10.1097/ACM.0000000000003712
16. Ledwos N, Mirchi N, Bissonnette V, Winkler-Schwartz A, Yilmaz R, Del Maestro RF. Virtual Reality Anterior Cervical Discectomy and Fusion Simulation on the Novel Sim-Ortho Platform: Validation Studies. *Oper Neurosurg (Hagerstown)*. 2020/12/15 2020;20(1):74-82. doi:10.1093/ons/opaa269
17. Delgado-Fernández J, Frade-Porto N, Blasco G, et al. Simulation with 3D Neuronavigation for Learning Cortical Bone Trajectory Screw Placement. *J Neurol Surg A Cent Eur Neurosurg*. 2020/12/1 2020;82(3):262-269. doi:10.1055/s-0040-1715485
18. Mendelis JR, Gomez JA, Lo Y, Hanstein R, Levy IM. Validity of a Novel Free-Hand Pedicle Screw Placement Training Tool. *J Am Acad Orthop Surg*. 2020/2/15 2020;28(4):e172-e180. doi:10.5435/JAAOS-D-18-00805

19. Martin JA, Regehr, R., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C., Brown, M. Objective structured assessment of technical skill (OSATS) for surgical residents. *British Journal of Surgery*. 1997;84:5.
20. Doyle JD, Webber EM, Sidhu RS. A universal global rating scale for the evaluation of technical skills in the operating room. *Am J Surg*. May 2007;193(5):551-5; discussion 555. doi:10.1016/j.amjsurg.2007.02.003
21. Vaidya A, Aydin A, Ridgley J, Raison N, Dasgupta P, Ahmed K. Current Status of Technical Skills Assessment Tools in Surgery: A Systematic Review. *J Surg Res*. Feb 2020;246:342-378. doi:10.1016/j.jss.2019.09.006
22. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Darzi A. Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. *Ann Surg*. Feb 2008;247(2):372-9. doi:10.1097/SLA.0b013e318160b371
23. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg*. Jul 2005;190(1):107-13. doi:10.1016/j.amjsurg.2005.04.004
24. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. Oct 10 2013;369(15):1434-42. doi:10.1056/NEJMsa1300625
25. Palter VN, Grantcharov TP. A prospective study demonstrating the reliability and validity of two procedure-specific evaluation tools to assess operative competence in laparoscopic colorectal surgery. *Surg Endosc*. Sep 2012;26(9):2489-503. doi:10.1007/s00464-012-2220-9
26. Larsen CR, Grantcharov T, Schouenborg L, Ottosen C, Soerensen JL, Ottesen B. Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. *BJOG*. Jun 2008;115(7):908-16. doi:10.1111/j.1471-0528.2008.01732.x
27. Cao CG, MacKenzie CL, Ibbotson JA, Turner LJ, Blair NP, Nagy AG. Hierarchical decomposition of laparoscopic procedures. *Stud Health Technol Inform*. 1999;62:83-9.
28. Dath D, Regehr,G. , Birch, D., Schlachta, C.,Poulin, E., Mamazza,J., et al. . Toward reliable operative assessment: the reliability and feasibility of videotaped assessment of laparoscopic technical skills. . *Surgical Endoscopy and Other Interventional Techniques*. 2004;18(12):4.
29. Sarker SK, Albrani T, Zaman A, Patel B. Procedural performance in gastrointestinal endoscopy: An assessment and self-appraisal tool. *The American Journal of Surgery*. 2008/09/01/ 2008;196(3):450-455. doi:https://doi.org/10.1016/j.amjsurg.2007.10.024
30. Sirimanna P, Boyce S, Gunanayagam P, Gladman MA, Naganathan V. Development of a rating scale for objective assessment of performance in laparoscopic appendicectomy surgery. *ANZ J Surg*. Jul 2022;92(7-8):1724-1730. doi:10.1111/ans.17601
31. Grotle M. SMC, Fjeld O., Grøvle L., Helgeland J., Storheim K. Lumbar spine surgery across 15 years: trends, complications and reoperations in a longitudinal observational study from Norway [Internet]. *BMJ Open*. 2019;9 e028743.
32. Steele L, Mukherjee S, Stratton-Powell A, Anderson I, Timothy J. Extent of medicolegal burden in neurosurgery - An analysis of the National Health Service Litigation Authority Database. *Br J Neurosurg*. 2015;29(5):622-9. doi:10.3109/02688697.2015.1054362
33. Quraishi NA, Hammett TC, Todd DB, Bhutta MA, Kapoor V. Malpractice litigation and the spine: the NHS perspective on 235 successful claims in England. *Eur Spine J*. May 2012;21 Suppl 2(Suppl 2):S196-9. doi:10.1007/s00586-012-2203-5
34. Street JT, Lenehan BJ, DiPaola CP, et al. Morbidity and mortality of major adult spinal surgery. A prospective cohort analysis of 942 consecutive patients. *Spine J*. Jan 2012;12(1):22-34. doi:10.1016/j.spinee.2011.12.003
35. Madsbu MA, Salvesen O, Carlsen SM, et al. Surgery for herniated lumbar disc in private vs public hospitals: A pragmatic comparative effectiveness study. *Acta Neurochir (Wien)*. Mar 2020;162(3):703-711. doi:10.1007/s00701-019-04195-7

36. Stienen MN, Smoll NR, Hildebrandt G, Schaller K, Gautschi OP. Early surgical education of residents is safe for microscopic lumbar disc surgery. *Acta Neurochir (Wien)*. Jun 2014;156(6):1205-14. doi:10.1007/s00701-014-2070-4
37. Aggarwal R, Mytton OT, Derbrew M, et al. Training and simulation for patient safety. *Qual Saf Health Care*. Aug 2010;19 Suppl 2:i34-43. doi:10.1136/qshc.2009.038562
38. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009/7/21 2009;339:b2535. doi:10.1136/bmj.b2535
39. Boody BS, Hashmi SZ, Rosenthal BD, et al. The Effectiveness of Bioskills Training for Simulated Lumbar Pedicle Screw Placement. *Global Spine J*. Sep 2018;8(6):557-562. doi:10.1177/2192568217743505
40. Sundar SJ, Healy AT, Kshettry VR, Mroz TE, Schlenk R, Benzel EC. A pilot study of the utility of a laboratory-based spinal fixation training program for neurosurgical residents. *J Neurosurg Spine*. 2016/5 2016;24(5):850-856. doi:10.3171/2015.8.SPINE15119
41. Tanner G, Vojdani S, Komatsu DE, Barsi JM. Development of a Saw Bones Model for training pedicle screw placement in scoliosis. *BMC Res Notes*. 2017/12/4 2017;10(1):678. doi:10.1186/s13104-017-3029-3
42. Gottschalk MB, Yoon ST, Park DK, Rhee JM, Mitchell PM. Surgical training using three-dimensional simulation in placement of cervical lateral mass screws: a blinded randomized control trial. *Spine J*. 2015/1/1 2015;15(1):168-175. doi:10.1016/j.spinee.2014.08.444
43. Woodrow SI, Dubrowski A, Khokhotva M, Backstein D, Rampersaud YR, Massicotte EM. Training and evaluating spinal surgeons: the development of novel performance measures. *Spine (Phila Pa 1976)*. Dec 1 2007;32(25):2921-5. doi:10.1097/BRS.0b013e31815b6495
44. Boody BS, Rosenthal BD, Jenkins TJ, Patel AA, Savage JW, Hsu WK. The Effectiveness of Bioskills Training for Simulated Open Lumbar Laminectomy. *Global Spine J*. 2017/6/23 2017;7(8):794-800. doi:10.1177/2192568217703337
45. Ghobrial GM, Balsara K, Maulucci CM, et al. Simulation Training Curricula for Neurosurgical Residents: Cervical Foraminotomy and Durotomy Repair Modules. *World Neurosurg*. 2015/9 2015;84(3):751-5.e1-7. doi:10.1016/j.wneu.2015.04.056
46. Garcia J, Yang Z, Mongrain R, Leask RL, Lachapelle K. 3D printing materials and their use in medical education: a review of current technology and trends for the future. *BMJ Simul Technol Enhanc Learn*. 2018/1 2018;4(1):27-40. doi:10.1136/bmjstel-2017-000234
47. Bow H, Zuckerman SL, Griffith B, et al. A 3D-Printed Simulator and Teaching Module for Placing S2-Alar-Iliac Screws. *Oper Neurosurg (Hagerstown)*. 2020/3/1 2020;18(3):339-346. doi:10.1093/ons/opz161
48. Park HJ, Wang C, Choi KH, Kim HN. Use of a life-size three-dimensional-printed spine model for pedicle screw instrumentation training. *J Orthop Surg Res*. 2018/4/16 2018;13(1):86. doi:10.1186/s13018-018-0788-z
49. Adermann J, Geissler N, Bernal LE, Kotzsch S, Korb W. Development and validation of an artificial wetlab training system for the lumbar discectomy. *Eur Spine J*. 2014/9 2014;23(9):1978-1983. doi:10.1007/s00586-014-3257-3
50. Bhatia N, Palispis WA, Urakov T, et al. Establishing validity of the fundamentals of spinal surgery (FOSS) simulator as a teaching tool for orthopedic and neurosurgical trainees. *Spine J*. 2020/4 2020;20(4):580-589. doi:10.1016/j.spinee.2019.11.008
51. Gardeck AM, Pu X, Yang Q, Polly DW, Jones KE. The effect of simulation training on resident proficiency in thoracolumbar pedicle screw placement using computer-assisted navigation. *J Neurosurg Spine*. 2020/9/4 2020:1-8. doi:10.3171/2020.5.SPINE2067
52. Ray WZ, Ganju A, Harrop JS, Hoh DJ. Developing an anterior cervical diskectomy and fusion simulator for neurosurgical resident training. *Neurosurgery*. 2013/10 2013;73 Suppl 1:100-106. doi:10.1227/NEU.0000000000000088

53. Walker JB, Perkins E, Harkey HL. A novel simulation model for minimally invasive spine surgery. *Neurosurgery*. 2009/12 2009;65(6 Suppl):188-95; discussion 195. doi:10.1227/01.NEU.0000341534.82210.1B
54. Ghobrial GM, Anderson PA, Chitale R, Campbell PG, Lobel DA, Harrop J. Simulated spinal cerebrospinal fluid leak repair: an educational model with didactic and technical components. *Neurosurgery*. 2013/10 2013;73 Suppl 1:111-115. doi:10.1227/NEU.0000000000000100
55. Ferguson D, Agyemang K, Barrett C, Mathieson C. A low cost dural closure simulation model for tomorrow's spinal neurosurgeons. *Br J Neurosurg*. 2019/5/4 2019;33(3):337-340. doi:10.1080/02688697.2018.1540765
56. Harrop J, Rezai AR, Hoh DJ, Ghobrial GM, Sharan A. Neurosurgical training with a novel cervical spine simulator: posterior foraminotomy and laminectomy. *Neurosurgery*. 2013/10 2013;73 Suppl 1:94-99. doi:10.1227/NEU.0000000000000103
57. Chitale R, Ghobrial GM, Lobel D, Harrop J. Simulated lumbar minimally invasive surgery educational model with didactic and technical components. *Neurosurgery*. 2013/10 2013;73 Suppl 1:107-110. doi:10.1227/NEU.0000000000000091
58. Podolsky DJ, Martin AR, Whyne CM, Massicotte EM, Hardisty MR, Ginsberg HJ. Exploring the role of 3-dimensional simulation in surgical training: feedback from a pilot study. *J Spinal Disord Tech*. 2010/12 2010;23(8):e70-4. doi:10.1097/BSD.0b013e3181d345cb
59. Tortolani PJ, Moatz BW, Parks BG, Cunningham BW, Sefter J, Kretzer RM. Cadaver training module for teaching thoracic pedicle screw placement to residents. *Orthopedics*. 2013/9 2013;36(9):e1128-33. doi:10.3928/01477447-20130821-13
60. Buchanan IA, Min E, Pham MH, et al. Simulation of Dural Repair in Minimally Invasive Spine Surgery With the Use of a Perfusion-Based Cadaveric Model. *Oper Neurosurg (Hagerstown)*. 2019/12/1 2019;17(6):616-621. doi:10.1093/ons/opz041
61. Calio BP, Kepler CK, Koerner JD, Rihn JA, Millhouse P, Radcliff KE. Outcome of a Resident Spine Surgical Skills Training Program. *Clin Spine Surg*. 2017/10 2017;30(8):E1126-E1129. doi:10.1097/BSD.0000000000000211
62. Hou Y, Lin Y, Shi J, Chen H, Yuan W. Effectiveness of the Thoracic Pedicle Screw Placement Using the Virtual Surgical Training System: A Cadaver Study. *Oper Neurosurg (Hagerstown)*. 2018/12/1 2018;15(6):677-685. doi:10.1093/ons/opy030
63. 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*. 2018/6 2018;138(6):777-782. doi:10.1007/s00402-018-2906-0
64. Chen T, Zhang Y, Ding C, et al. Virtual reality as a learning tool in spinal anatomy and surgical techniques. *Spine J*. 2021/6 2021;6:100063. doi:10.1016/j.xnsj.2021.100063
65. Luca A, Giorgino R, Gesualdo L, et al. Innovative Educational Pathways in Spine Surgery: Advanced Virtual Reality-Based Training. *World Neurosurg*. 2020/8 2020;140:674-680. doi:10.1016/j.wneu.2020.04.102
66. Luciano CJ, Banerjee PP, Bellotte B, et al. Learning Retention of Thoracic Pedicle Screw Placement Using a High-Resolution Augmented Reality Simulator With Haptic Feedback. *Operative Neurosurgery*. 2011/9 2011;69:ons14. doi:10.1227/NEU.0b013e31821954ed
67. Luciano CJ, Banerjee PP, Sorenson JM, et al. Percutaneous Spinal Fixation Simulation With Virtual Reality and Haptics. *Neurosurgery*. 2013/1 2013;72:A89. doi:10.1227/NEU.0b013e3182750a8d
68. 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
69. Gasco J, Patel A, Ortega-Barnett J, et al. Virtual reality spine surgery simulation: an empirical study of its usefulness. *Neurol Res*. 2014/11 2014;36(11):968-973. doi:10.1179/1743132814Y.0000000388

70. Gasco J, Holbrook TJ, Patel A, et al. Neurosurgery simulation in residency training: feasibility, cost, and educational benefit. *Neurosurgery*. 2013/10 2013;73 Suppl 1:39-45. doi:10.1227/NEU.0000000000000102
71. Van Le C, Tromp JG, Puri V. Using 3D simulation in medical education: A comparative test of teaching anatomy using virtual reality. 2018/10 2018:21-33. doi:10.1002/9781119509875.ch2
72. Bohl MA, Mooney MA, Repp GJ, et al. The Barrow Biomimetic Spine: Fluoroscopic Analysis of a Synthetic Spine Model Made of Variable 3D-printed Materials and Print Parameters. *Spine*. 2018/12/1 2018;43(23):E1368-E1375. doi:10.1097/BRS.0000000000002715
73. Bova FJ, Rajon DA, Friedman WA, et al. Mixed-reality simulation for neurosurgical procedures. *Neurosurgery*. 2013/10 2013;73 Suppl 1:138-145. doi:10.1227/NEU.0000000000000113
74. Eftekhar B, Ghodsi M, Ketabchi E, Rasaee S. Surgical simulation software for insertion of pedicle screws. *Neurosurgery*. 2002/1 2002;50(1):222-3; discussion 223-4. doi:10.1097/00006123-200201000-00038
75. Aoude A, Alhamzah H, Fortin M, Jarzem P, Ouellet J, Weber MH. The use of computer-assisted surgery as an educational tool for the training of orthopedic surgery residents in pedicle screw placement: a pilot study and survey among orthopedic residents. *Can J Surg*. 2016/12 2016;59(6):391-398. doi:10.1503/cjs.012115
76. Liu G-J, Zhang S-X, Qiu M-G, Tan L-W, Li Q-Y, Li K. A novel technique for three-dimensional reconstruction for surgical simulation around the craniocervical junction region. *Int Surg*. 2011/7 2011;96(3):274-280. doi:10.9738/cc14.1
77. Shore EM, Lefebvre GG, Grantcharov TP. Gynecology resident laparoscopy training: present and future. *Am J Obstet Gynecol*. Mar 2015;212(3):298-301, 298 e1. doi:10.1016/j.ajog.2014.07.039
78. Sirimanna P, Gladman MA. Development of a proficiency-based virtual reality simulation training curriculum for laparoscopic appendicectomy. *ANZ J Surg*. Oct 2017;87(10):760-766. doi:10.1111/ans.14135
79. Stride HP, George BC, Williams RG, et al. Relationship of procedural numbers with meaningful procedural autonomy in general surgery residents. *Surgery*. Mar 2018;163(3):488-494. doi:10.1016/j.surg.2017.10.011
80. Abdelrahman T, Long J, Egan R, Lewis WG. Operative Experience vs. Competence: A Curriculum Concordance and Learning Curve Analysis. *J Surg Educ*. Jul-Aug 2016;73(4):694-8. doi:10.1016/j.jsurg.2016.01.011
81. Macnamara BN, Moreau D, Hambrick DZ. The Relationship Between Deliberate Practice and Performance in Sports: A Meta-Analysis. *Perspect Psychol Sci*. May 2016;11(3):333-50. doi:10.1177/1745691616635591
82. Shore EM, Lefebvre GG, Husslein H, Bjerrum F, Sorensen JL, Grantcharov TP. Designing a Standardized Laparoscopy Curriculum for Gynecology Residents: A Delphi Approach. *J Grad Med Educ*. Jun 2015;7(2):197-202. doi:10.4300/JGME-D-14-00548.1
83. Dalkey N, Helmer, O. An Experimental Application of the Delphi Method to the Use of Experts. *Management Science*. 1963;9:458-467.
84. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs*. Oct 2000;32(4):1008-15.
85. Linstone H, Turoff M. *The Delphi Method: Techniques and Applications*. vol 18. 1975.
86. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*. 2004/12/01/ 2004;42(1):15-29. doi:https://doi.org/10.1016/j.im.2003.11.002
87. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. *J Clin Epidemiol*. Dec 2003;56(12):1150-6. doi:10.1016/s0895-4356(03)00211-7
88. Burns KEA, Duffett M, Kho ME, et al. A guide for the design and conduct of self-administered surveys of clinicians. *Canadian Medical Association Journal*. 2008;179(3):245-252. doi:10.1503/cmaj.080372

89. Long DM. Competency-based Residency Training: The Next Advance in Graduate Medical Education. *Academic Medicine*. 2000;75(12):6.
90. Drummond KJ, Hunn BHM, McAlpine HE, Jones JJ, Davies MA, Gull S. Challenges in the Australasian neurosurgery training program: who should be trained and where should they train? *Neurosurg Focus*. Mar 1 2020;48(3):E10. doi:10.3171/2019.12.FOCUS19870
91. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. *Am J Surg*. Oct 2011;202(4):469-480 e6. doi:10.1016/j.amjsurg.2010.10.020
92. Kramp KH, van Det MJ, Veeger NJ, Pierie JP. Validity, reliability and support for implementation of independence-scaled procedural assessment in laparoscopic surgery. *Surg Endosc*. Jun 2016;30(6):2288-300. doi:10.1007/s00464-015-4254-2
93. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. Jun 2016;15(2):155-63. doi:10.1016/j.jcm.2016.02.012
94. Parker SL, Xu R, McGirt MJ, Witham TF, Long DM, Bydon A. Long-term back pain after a single-level discectomy for radiculopathy: incidence and health care cost analysis. *J Neurosurg Spine*. Feb 2010;12(2):178-82. doi:10.3171/2009.9.SPINE09410
95. Ashraf H, Gunda D, Morgan FH, et al. Impact of work hour restrictions on the operative experience of general surgical residents: A systematic review. *Surgery in Practice and Science*. 2023;15doi:10.1016/j.sipas.2023.100222
96. Ferguson PC, Caverzagie KJ, Nousiainen MT, Snell L, Collaborators I. Changing the culture of medical training: An important step toward the implementation of competency-based medical education. *Med Teach*. Jun 2017;39(6):599-602. doi:10.1080/0142159X.2017.1315079
97. Bhattacharya S. Competency-based medical education: An overview. *Annals of Medical Science & Research*. 2023;2(3):132-138. doi:10.4103/amsr.amsr_27_23
98. Campbell C, Crebbin, W., Hickey, K., Stokes, M-L., Watters, D. Work-based assessment: a practical guide. Pamphlet. Tripartite Alliance Work-based Assessment Workshops series. Accessed 8 November 2024, 2024. https://www.surgeons.org/-/media/Project/RACS/surgeons-org/files/becoming-a-surgeon-trainees/work-based-assessment-a-practical-guide.pdf?rev=64c62242e777411eb43be8ac781dfa4a&hash=DCEE633AC11B7EE63975DF1A6948C99A
99. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. Sep 1990;65(9 Suppl):S63-7. doi:10.1097/00001888-199009000-00045
100. Aryal K, Hamed M, Currow C. The usefulness of work-based assessments in higher surgical training: A systematic review. *Int J Surg*. Oct 2021;94:106127. doi:10.1016/j.ijsu.2021.106127
101. Shalhoub J, Marshall DC, Ippolito K. Perspectives on procedure-based assessments: a thematic analysis of semistructured interviews with 10 UK surgical trainees. *BMJ Open*. Mar 24 2017;7(3):e013417. doi:10.1136/bmjopen-2016-013417
102. Gaunt A, Patel A, Royle J, et al. What do surgeons and trainees think of WBAs and how do they use them? *The Bulletin of the Royal College of Surgeons of England*. 2016;98(9):408-414. doi:10.1308/rcsbull.2016.408
103. Rela M, Price T. Review of the validity of DOPS as an assessment tool for the procedural skills of surgical trainees. *Ann R Coll Surg Engl*. Sep 2023;105(7):599-606. doi:10.1308/rcsann.2022.0052
104. Lewis T. Luigi Mangione Said He Had Spondylolisthesis: What Is It? Accessed 11 December 2024, 2024. https://www.scientificamerican.com/article/luigi-mangione-said-he-had-spondylolisthesis-what-is-it/
105. Stulberg JJ, Huang R, Kreutzer L, et al. Association Between Surgeon Technical Skills and Patient Outcomes. *JAMA Surgery*. 2020;155(10):960-968. doi:10.1001/jamasurg.2020.3007
106. Bisgaard CH, Rubak SLM, Rodt SA, Petersen JAK, Musaeus P. The effects of graduate competency-based education and mastery learning on patient care and return on investment: a narrative review of basic anesthetic procedures. *BMC Med Educ*. Jun 28 2018;18(1):154. doi:10.1186/s12909-018-1262-7

107. Meling TR, Meling TR. The impact of surgical simulation on patient outcomes: a systematic review and meta-analysis. *Neurosurg Rev*. Apr 2021;44(2):843-854. doi:10.1007/s10143-020-01314-2
108. Cheng H, Clymer JW, Po-Han Chen B, et al. Prolonged operative duration is associated with complications: a systematic review and meta-analysis. *J Surg Res*. Sep 2018;229:134-144. doi:10.1016/j.jss.2018.03.022
109. Aydin A, Ahmed K, Abe T, et al. Effect of Simulation-based Training on Surgical Proficiency and Patient Outcomes: A Randomised Controlled Clinical and Educational Trial. *Eur Urol*. Apr 2022;81(4):385-393. doi:10.1016/j.eururo.2021.10.030