Education and Training in Ultrasound-Guided Regional Anaesthesia

A thesis submitted for the fulfilment of the degree of

Doctor of Philosophy

by

Dr Alwin Hooi Hock CHUAN

M.B,B.S. (Hons Class I) PGCertCU F.A.N.Z.C.A.

Faculty of Medicine and Health Sciences

Macquarie University

Sydney, Australia

July 2016

Field of research code: 110301 (Anaesthesiology)

Abstract

The concept of using Doppler ultrasound to guide regional anaesthesia performance was first described in 1978. It was, however, the introduction into clinical practice in the last 15 years of portable, affordable, high resolution, bedside ultrasound machines that has revolutionised the practice of regional anaesthesia. Visualising anatomical structures, and guiding a needle to target structures under direct ultrasound guidance, is now considered best practice.

Ultrasound-guided regional anaesthesia, compared to traditional nerve localisation techniques such as nerve stimulation and paraesthesia, has been shown to improve efficacy and efficiency, and reduce the risk of local anaesthesia systemic toxicity and pneumothorax. Ultrasound has allowed the introduction of novel approaches to thoracoabdominal and neuraxial blocks. This imaging technique is increasingly being used in pain medicine, complementing and in some instances replacing, the image intensifier and computed tomography-guided interventional procedures.

In contrast, novices attempting ultrasound guided regional anaesthesia exhibit suboptimal behaviours, including visual-spatial disorientation, rigid procedural thinking, and needle manipulation without confirmation of positioning. The root problem is that teaching of regional anaesthesia is variable in quality and is non-systematic.

The reasons are complex, but include variability in supervision, worsening production pressures in busy tertiary hospitals curtailing time for teaching, shortening of trainee training times, and resistance by clinicians for new techniques. Compounding these problems was a lack of validated, reliable and objective tools to assess ultrasound-guided regional anaesthesia performance.

This is addressed in the first four studies of this thesis. Studies 1 and 2 evaluated the psychometric properties of the direct observation of procedural skills assessment tool used in the current training curriculum of the Australian and New Zealand College of Anaesthetists. I found that inter-assessor reliability is poor, which has important consequences as this tool is used for trainee assessment and structured feedback. Study 3 evaluated a checklist and global rating scale designed specifically for ultrasound-guided regional anaesthesia. This tool showed good construct validity, and that a deconstructed, itemised checklist is useful for teaching complex skills such as regional anaesthesia.

Study 4 described the design, creation, and validation of the Regional Anaesthesia Procedural Skills (RAPS) assessment tool. RAPS has evidence for face validity, construct

validity, test-retest reliability, external reliability, and feasibility as an assessment tool for all regional anaesthesia blocks, including ultrasound-guided techniques. The RAPS tool can thus be used for clinical assessment of trainees, as well as a reliable measure of performance in participants in education research.

The next two studies investigated factors by which training in ultrasound-guided regional anaesthesia can be improved. Study 5 was a randomised controlled trial comparing whether fresh-frozen human cadavers were superior to meat-based models for teaching ultrasound-guided regional anaesthesia. I found that while face validity and qualitative satisfaction was superior for cadavers, there was no quantitative difference in efficacy, efficiency or errors committed in a part-task technical skills test.

Study 6 was an exploratory study in whether visuospatial ability influences sonography performance. In novices performing brachial plexus sonography and reliant only on discovery learning, three visuospatial factors were found to be influential: spatial visualisation, spatial relations, and speed of closure. The standardised visuospatial test battery can thus identify novices who will likely struggle with sonography. This opens an avenue for training tailored to an individual's strengths and weaknesses.

Declaration

This is to certify that:

(i) this thesis comprises only original work completed by the author for degree of Doctor of Philosophy at Macquarie University, except as indicated in the Statement of Co-authorship and Acknowledgments;

(ii) the thesis contains only material that has not been published by any other person, except where appropriate references and acknowledgments have been made in the text;

(iii) the thesis has not been previously submitted for a higher degree in another university;

(iv) all studies included in this thesis by publication have been approved by a properly formed and suitable human research ethics committee:

- Study 1 - *Psychometric Evaluation of a Direct Observation of Procedural Skills Assessment Tool for Ultrasound-Guided Regional Anaesthesia*
  Approved by Macquarie University HREC (Human Sciences), 27 Aug 2013 (Ref: 5201300089) and the St Vincent's Hospital (Melbourne) Research Governance Unit, 18 Sept 2012 (Ref: HREC-A/LRR 134/12)
- Study 2 - *Reliability of the Direct Observation of Procedural Skills Assessment Tool for Ultrasound-Guided Regional Anaesthesia*
  Approved by Macquarie University HREC (Human Sciences), 21 Nov 2013 (Ref: 5201300757)
- Study 3 - *Evaluation of a Task-specific Checklist and Global Rating Scale for Ultrasound-Guided Regional Anaesthesia*
  Approved by Macquarie University HREC (Human Sciences), 27 Aug 2013 (Ref: 5201300089) and the St Vincent's Hospital (Melbourne) Research Governance Unit, 18 Sept 2012 (Ref: HREC-A/LRR 134/12)
- Study 4 - *Design and Validation of the Regional Anaesthesia Procedural Skills (RAPS) Assessment Tool*
  Approved by South Western Sydney Local Health District HREC, 19 May 2014 (Ref: HREC/11/LPOOL/387) and the St Vincent's Hospital (Melbourne) Research Governance Unit, 22 April 2014 (Ref: HREC-A/LRR 134/12)

- Study 5 - *A Randomised Controlled Trial Comparing Meat-based with Human Cadaveric Models for Teaching Ultrasound-Guided Regional Anaesthesia*
  Approved by Macquarie University HREC (Human Sciences), 24 July 2014
  (Ref: 5201400737)
- Study 6 - *Visuospatial Ability and Novice Brachial Plexus Sonography Performance*
  Approved by South Western Sydney Local Health District HREC, 16 Oct 2014
  (Ref: HREC/14/LPOOL/394)

and

(v) the thesis is less than 100,000 words in length, exclusive of tables, figure legends, references, and appendices.

Statement of Co-authorship

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

- Study 1 - *Psychometric Evaluation of a Direct Observation of Procedural Skills Assessment Tool for Ultrasound-Guided Regional Anaesthesia*
  The candidate was involved with original protocol design, formulation of the trainee interviews, pilot assessment, and blinded assessor. All authors assisted with manuscript preparation. Specific authors' contributions were:
  A/Prof Michael Barrington, University of Melbourne, Melbourne
  Protocol design, trainee interview design
  Dr Damien Castanelli, Monash University, Melbourne
  Dr Michael Herrick, Geisel School of Medicine, Dartmouth, USA
  Dr Andrew Lansdown, Royal Prince Alfred Hospital, Sydney
  Dr Lisa Lin, Austin Hospital, Melbourne
  Dr Irene Ng, Royal Melbourne Hospital, Melbourne
  Above authors were involved with pilot assessment, blinded assessors
  Dr Roman Kluger, St Vincent's Hospital, Melbourne
  Statistical analysis
  Mr Matthew Watson, St Vincent's Hospital, Melbourne
  Data collection
  Dr Daniel Wong, St Vincent's Hospital, Melbourne
  Protocol design, trainee interview design

- Study 2 - *Reliability of the Direct Observation of Procedural Skills Assessment Tool for Ultrasound-Guided Regional Anaesthesia*
  The candidate was primary author, original protocol design, recruited, and analysed the data in collaboration with:
  A/Professor Michael Barrington, University of Melbourne, Melbourne
  Dr Petra Graham, Macquarie University, Sydney
  Professor Brian Jolly, University of Newcastle, Newcastle, Australia
  Dr Natalie Smith, University of Wollongong, Australia
  Dr Sobana Thillainathan, Liverpool Hospital, Sydney
  Dr Daniel Wong, St Vincent's Hospital, Melbourne

- Study 3 - *Evaluation of a Task-specific Checklist and Global Rating Scale for Ultrasound-Guided Regional Anaesthesia*

  The candidate was involved with original protocol design, formulation of the trainee interviews, pilot assessment, and blinded assessor. All authors assisted with manuscript preparation. Specific authors' contributions were:

  A/Prof Michael Barrington, University of Melbourne, Melbourne

  Protocol design, trainee interview design

  Dr Damien Castanelli, Monash University, Melbourne

  Dr Michael Herrick, Geisel School of Medicine, Dartmouth, USA

  Dr Andrew Lansdown, Royal Prince Alfred Hospital, Sydney

  Dr Lisa Lin, Austin Hospital, Melbourne

  Dr Irene Ng, Royal Melbourne Hospital, Melbourne

  Above authors were involved with pilot assessment, blinded assessors

  Dr Roman Kluger, St Vincent's Hospital, Melbourne

  Statistical analysis

  Mr Matthew Watson, St Vincent's Hospital, Melbourne

  Data collection

  Dr Daniel Wong, St Vincent's Hospital, Melbourne

  Protocol design, trainee interview design

- Study 4 - *Design and Validation of the Regional Anaesthesia Procedural Skills (RAPS) Assessment Tool*

  The candidate was primary author, original protocol design, recruited, and analysed the data in collaboration with:

  Dr David Auyong, Virginia Mason Medical Centre, Seattle, USA

  A/Prof Michael Barrington, University of Melbourne, Melbourne

  Dr Andrew Cameron, Middlemore Hospital, Auckland, New Zealand

  Prof Kirsty Forrest, Macquarie University, Sydney

  Dr Biljana Germanoska, Liverpool Hospital, Sydney

  Dr Petra Graham, Macquarie University, Sydney

  Dr Yean Chin Lim, Changi General Hospital, Singapore

  Ms Linda Pope, Liverpool Hospital, Sydney

  Prof Colin Royse, University of Melbourne, Melbourne

  Dr Daniel Wong, St Vincent's Hospital, Melbourne

- Study 5 - *A Randomised Controlled Trial Comparing Meat-based with Human Cadaveric Models for Teaching Ultrasound-Guided Regional Anaesthesia*
  The candidate was primary author, original protocol design, recruited, and analysed the data in collaboration with:
  Dr Harmeet Aneja, Sir Charles Gairdner Hospital, Perth
  A/Prof Richard Appleyard, Macquarie University, Sydney
  Mr Nicholas Duce, Liverpool Hospital, Sydney
  Prof Kirsty Forrest, Macquarie University, Sydney
  Dr Yean Chin Lim, Changi General Hospital, Singapore
  Prof Colin Royse, University of Melbourne, Melbourne

- Study 6 - *Visuospatial Ability and Novice Brachial Plexus Sonography Performance*
  The candidate was primary author, original protocol design, recruited, and analysed the data in collaboration with:
  Mr Joseph Descallar, Ingham Institute of Applied Medical Research, Sydney
  Mr Nicholas Duce, Liverpool Hospital, Sydney
  Dr Lauren Gillett, Ingham Institute of Applied Medical Research, Sydney
  Dr Steven Siu, Liverpool Hospital, Sydney
  Dr Minh T Tran, Liverpool Hospital, Sydney

Acknowledgments

I am indebted to many people who have helped with this thesis. Firstly, to my wife Cindy, who made (and continues to make) many sacrifices whilst I run around endlessly from one project to another. You provide the stability and anchor for Natasha and Samuel, and without your support this work would not have been possible.

An enormous gratitude also goes to Professor Colin Royse. It was Colin who first encouraged me to consider, and then supported me, in undertaking this thesis. Under his mentorship, constant guidance, and personal counsel, I have learnt about research, academic vigour, yet also work, family and life. This was always done with good cheer and clarity. It has been a privilege and honour to be one of your students.

To Professors Kirsty Forrest and Mark Connor, many thanks for being my supervisors and always ready with a valuable answer, insight, or lending a patient ear to listen to my ideas. You have both been great supervisors and supported me over the past 4 years.

I have been blessed with many friendships and been inspired by many colleagues throughout my career. Singling out individuals is always fraught with difficulty. To Ajay Kumar, thank you for the mentorship that a young registrar needed, and for introducing me to ultrasound. To your credit you allowed me to explore the possibilities of perioperative ultrasonography and helped me get started on this journey. To Ammar Beck and Tillman Boesel, those six months I spent with you was a turning point in giving me the self-confidence to grow into a career in anaesthesia. Thanks for the many laughs and truths that we shared. To David M. Scott, I could not ask for a better friend and collaborator in writing the textbook on ultrasound-guided regional anaesthesia. Here is to more bluegrass jazz, good pinot noir, and good company in the future.

To my colleagues in Liverpool Hospital, I thank you for your constant and much appreciated support to establish a research unit within the department of anaesthesia. Instrumental is Carlo Vernier as head of department, as well as members of the ultrasound group: Tung Bui, Michael Ehrlich, Andrew Hill, BJ Huh, Fred Lee, Steven Siu, Clement Tiong, Minh T Tran, and Chris Wong. To the wider network of enthusiasts and friends, I thank Malcolm Albany, Harmeet Aneja, Phil Cowlishaw, Clement Fong, Peter Hebbard, Andrew Lansdown, Chris Mitchell, Paul Soeding, and Daniel Wong for your willingness to share your time and skills with the numerous studies and projects. Perhaps even more importantly, for sharing your lives outside of medicine.

This candidature was financially supported by the National Health and Medical Research Council Postgraduate Fellowship 2013 - 2015, and by the Australian Society of Anaesthetists PhD Support Grant 2013. I am grateful for their trust and confidence in providing me a start to my academic endeavours.

Dr Alwin Chuan

July 2016

Publications, presentations and grants arising from thesis

Published papers

Study 1 - Watson MJ, Wong DM, Kluger R, **Chuan A**, Herrick MD, Ng I, Castanelli DJ, Lin L, Lansdown A, Barrington MJ. Psychometric evaluation of a direct observation of procedural skills assessment tool for ultrasound-guided regional anaesthesia. *Anaesthesia* 2014; 69**:**604-12.

Study 2 - **Chuan A**, Thillainathan S, Graham PL, Jolly B, Wong DM, Smith NA, Barrington MJ. Reliability of the Direct Observation of Procedural Skills Assessment Tool for Ultrasound-Guided Regional Anaesthesia. *Anaesth Intensive Care* 2016; 44(2):201-209.

Study 3 - Wong DM, Watson MJ, Kluger R, **Chuan A**, Herrick MD, Ng I, Castanelli DJ, Lin L, Lansdown A, Barrington MJ. Evaluation of a task-specific checklist and global rating scale for ultrasound-guided regional anesthesia. *Reg Anesth Pain Med* 2014; 39**:**399-408.

Study 4 - **Chuan A**, Graham PL, Wong DM, Barrington MJ, Auyong DB, Cameron AJD, Lim YC, Pope L, Germanoska B, Forrest K, Royse CF. Design and validation of the Regional Anaesthesia Procedural Skills Assessment Tool. *Anaesthesia* 2015; 70**:**1401-11.

Study 5 - **Chuan A**, Lim YC, Aneja H, Duce DA, Appleyard R, Forrest K, Royse CF. A Randomised Controlled Trial Comparing Meat-based with Human Cadaveric Models for Teaching Ultrasound-Guided Regional Anaesthesia. *Anaesthesia* 2016; Epub ahead of print DOI:10.1111/anae.13446

Study 6 - Duce NA, Gillett L, Descallar J, Tran MT, Siu SCM, **Chuan A**. Visuospatial Ability and Novice Brachial Plexus Sonography Performance. *Acta Anaesthesiologica Scandinavica* 2016; Epub ahead of print DOI:10.1111/aas.12757

Published Textbook

A Chuan, DM Scott. *Oxford Regional Anaesthesia: A Pocket Guide*. Oxford, UK. Oxford University Press, 2014. ISBN: 978-0-199-68423-6

Chapter contribution to Textbook

A Chuan, W Harrop-Griffiths. In: A Hadzic (ed). *Textbook of Regional Anesthesia and Acute Pain Management* 2nd Edition. New York, USA. McGraw-Hill Education, in production 2016.

Grants

1. National Health and Medical Research Council Postgraduate Fellowship (APP1056280)

2. Australian Society of Anaesthetists PhD Support Grant (2013)

Presentations at major national and international meetings

1. Invited speaker, Australian Society of Anaesthetists National Scientific Congress, Canberra 2013. "Assessment of Procedural Skills for Regional Anaesthesia: What Tools are Available?"

2. Invited speaker, Singapore Cadaveric Ultrasound-Guided Regional Anaesthesia conference, Singapore 2013. "Sonoanatomy and Lower Limb Blocks" and "Catheter Based Techniques using Ultrasound-Guided Regional Anaesthesia"

3. Invited speaker, World Congress of Regional Anaesthesia and Pain Therapy, Sydney 2013. "Monitoring Standards for Safety in Peripheral Nerve Blocks: Science and Practice"

4. Invited speaker, World Anaesthesia Congress, Bangkok 2013. "Safe Sedation in Regional Anaesthesia: When, Why, How?"

5. Invited speaker, Australian and New Zealand College of Anaesthetists Combined Scientific Meeting, Singapore 2014. "How I Approach Blocks Above the Clavicle: An Anatomical Approach"

6. Invited Speaker, Australasian Symposium on Ultrasound and Regional Anaesthesia, Auckland 2014. "Global Perspectives in Training and Education in Ultrasound-Guided Regional Anaesthesia - The Australasian View"

7. Invited Speaker, New York Society of Regional Anaesthesia - Kuala Lumpur 2014. "The Learning Curve for Ultrasound-Guided Regional Anaesthesia"

8. Invited Speaker, Australasian Symposium on Ultrasound and Regional Anaesthesia, Perth 2015. "What Should We Measure? Assessment Tools for Regional Anaesthesia"

9. Invited Speaker, British Society of Orthopaedic Anaesthetists Annual Scientific Meeting, Oxford University 2016. "Outcomes after Regional Anaesthesia for Orthopaedic Surgery"

10. Invited Speaker, Australian and New Zealand College of Anaesthetists Annual Scientific Meeting, Brisbane 2017. "Think Big: Big Ideas on Improving Training in Regional Anaesthesia"

Table of Contents

Table of figures

List of tables

Abbreviations

ABA	American Board of Anesthesiology
ABIM	American Board of Internal Medicine
ACGME	Accreditation Council for Graduate Medical Education
ADL	Activities of Daily Living
ANZCA	Australian and New Zealand College of Anaesthetists
ASRA	American Society of Regional Anesthesia and Pain Medicine
cusum	Cumulative sum
DOPS	Direct Observation of Procedural Skills
ESRA	European Society of Regional Anaesthesia and Pain Therapy
GA	General Anaesthesia
GRS	Global Rating Scale
ICC	Intraclass Correlation Coefficient
ICSAD	Imperial College Surgical Assessment Device
IRT	Item Response Theory
MCQ	Multiple Choice Question examinations
mini-CEX	mini-Clinical Evaluation eXercise
MSF	Multi-Source Feedback
OSATS	Objective Structured Assessment of Technical Skills
OSCE	Objective Structured Clinical Examination
PNS	Peripheral Nerve Stimulator
RA	Regional Anaesthesia
RAPS	Regional Anaesthesia Procedural Skills assessment tool
RCA	Royal College of Anaesthetists (United Kingdom)

STATS	Systemic Training and Assessment of Technical Skills
UGRA	Ultrasound-Guided Regional Anaesthesia
WBA	Workplace-based assessment
ZPD	Zone of Proximal Development

Preface

My introduction to peripheral regional anaesthesia was memorable and illustrative. As a young second year trainee in 2003, I was allocated to a vascular list with a senior consultant anaesthetist. Our next patient was an elderly man requiring an arteriovenous fistula for haemodialysis, with active tuberculosis on account of being immunosuppressed for his failing renal transplant. Being cognisant of this patient's multiple co-morbidities, I discussed aloud that an axillary brachial plexus block would be an ideal technique. Hoping to be taught - or at least be supervised - by my consultant, I was surprised to be told to head to the department library, read up on the technique, and perform the block unsupervised: she wasn't taught how to do nerve blocks during her training, and no-one else in the department (in a tertiary teaching hospital in the largest city in Australia) knew either. My first attempt to perform a peripheral nerve block was a spectacular failure. I found that this lack of education, training and exposure to regional anaesthesia was not uncommon.

By 2005, and nearing the end of my training, the concept of anaesthetists using ultrasound for clinical applications was rapidly gaining traction. It was serendipitous timing, as I was convinced that this technology had great potential, and I was looking for opportunities to gain skills in this new field. Supported by Dr Ajay Kumar, and under guidance from Professor Colin Royse and the Ultrasound Education Group, I embarked on a (what was then unique) perioperative ultrasound fellowship based on the triad of transthoracic echocardiography, ultrasound-guided vascular access, and ultrasound-guided regional anaesthesia. It was a formative experience, requiring hard work, and kindled an academic interest. Most importantly, I was put into contact with a national and international community of like-minded, dedicated and gifted anaesthesia colleagues and friends.

The road to the doctorate was circuitous. Major intermediate milestones were establishing a regional anaesthesia fellowship training program in a city not previously known as "block-friendly"; starting research activities in a city well known as "academic-lite"; convening ultrasound workshops and conferences; and starting a family. These all posed challenges, but were instructive and I gained many skills not written in textbooks.

In the meantime, regional anaesthesia underwent a renaissance as evidenced by the volume of journal articles, textbooks, curriculum changes, and popularity of workshops. Ultrasound machines improved with each generation, needles for use with ultrasound were designed, and nerve blocks made possible only with ultrasound-guidance were described. However, what has remained consistent is the inconsistency of training, and lack of quality control of the "end-product" after training. This has resulted in a wide range of skills in anaesthetists, and in some cases meant that our patients have been denied safe and effective regional anaesthesia.

In a prescient editorial for the British Journal of Anaesthesia written in 2005, Drs Nicholas Denny and William Harrop-Griffiths expressed their "one concern is that the successful use of ultrasound may demand more training, a more detailed knowledge of anatomy, and even greater manual dexterity, and that regional anaesthesia will therefore continue to the preserve of the committed few". These giants of regional anaesthesia themselves quote another doyen, Dr Alon Winnie: "When there are problems with any regional technique, look for the cause first on the proximal end of the needle".

As an enthusiast for ultrasound in anaesthesia, and a proponent for regional anaesthesia, my desire is that one day the problems at the proximal end of needles becomes a historical footnote. One way this goal is achieved is through high quality, evidence-based, training and educational interventions; hopefully this thesis will be a positive contribution to that goal.

Chapter 1

Introduction and Literature Review

1.1 Introduction

The use of ultrasound in regional anaesthesia procedures is a relatively recent addition to the skill set practiced by anaesthetists. Ultrasound-guided regional anaesthesia is an extensive and topical sub-specialty generating very active research, which continues to modify the evidence base. Teaching anaesthetists to be proficient in this new imaging modality is a priority, and further research is required to determine the best methods for education and training.

Section 1.2 of this literature review will describe a brief history of regional anaesthesia, and the adoption of ultrasound-guidance into the specialty. The specific technical and non-technical tasks required to safely perform ultrasound-guided regional anaesthesia will be summarised in section 1.3.

Section 1.4 is an overview of selected educational theories relevant to acquiring medical procedural skills. Sections 1.5 to 1.6 will review how trainers use assessment tools to evaluate and quantify the success of trainees in meeting educational outcomes. The importance of the psychometric properties of assessment tools will be discussed, as well as the three main statistical methods to evaluate reliability and validity.

Section 1.7 reviews the current tools used to score regional anaesthesia performance. Lastly, the hypotheses and aims of the seven individual studies that comprise this prospective doctorate is presented in section 1.8.

1.2 History of Regional Anaesthesia and Ultrasound-Guided Regional Anaesthesia

Regional anaesthesia is dependent on successful deposition of an active drug adjacent to nerve fibres to affect a sensory and/or motor conduction block of nerve impulse transmission. This aim thus requires a safe and reliable local anaesthetic drug, combined with a thorough knowledge of human anatomy and the ability to localise neural structures.

In the latter half of the 19th century, developments in technology, the pharmaceutical industry, and academic publications brought together elements of what is now recognisable as modern regional anaesthesia (RA). The first effective local anaesthetic agent was cocaine, although its anaesthetic properties were not appreciated for several decades (unlike its effects on the mental state), despite the alkaloid being isolated by 1859.[1] The first description of the clinical use of cocaine as a local anaesthetic occurred in 1884 by Karl Koller, who published his case report in The Lancet.[2] An ophthalmologist, Koller performed glaucoma surgery in an awake patient after administrating a topical solution of 3% cocaine. This breakthrough resulted in the popularisation of topical and subcutaneous infiltration anaesthesia for eye, ear, nose, and throat surgery. Over the subsequent decades, as the severe side effects, risk of overdose, and addictive qualities of cocaine became recognised, chemists were prompted to search for safer local anaesthetic agents.

This quickly occurred in the early 20th century, when a succession of ester-based local anaesthetic agents were produced: procaine, tetracaine, dibucaine, and later still, chloroprocaine. In the immediate period after World War 2, the amide-based compounds were manufactured, resulting in the drugs we continue to use today: lignocaine, bupivacaine, mepivacaine, and prilocaine. Further isolation and purification of the S-enantiomers have in turn allowed the commercialisation of ropivacaine and levobupivacaine.

In parallel with the introduction of local anaesthesia agents, equipment designed to deliver these drugs were developed. The hypodermic needle was independently invented by the Scottish physician Alexander Wood, and French surgeon Charles Pravaz in 1853. However, Wood published his findings first in 1855, describing the technique of injecting an active drug adjacent to nerves to cause sensory blockade in a case series.[3] Although the drug was morphine, and to treat neuralgia rather than a conduction block for anaesthesia or analgesia, Wood's paper nonetheless described the basic concept of RA.

1.2.1 Nerve localisation techniques in regional anaesthesia

Techniques for accurate placement of local anaesthesia adjacent to neural structures have greatly evolved. An ongoing challenge for any regional anaesthetist is to locate the target nerves reliably, and to guide the needle precisely to that target. In the first description of a brachial plexus block, performed by the American surgeon William Halsted in 1885, the target neural structure was located by direct vision after surgical dissection and the nerves were bathed in a cocaine solution.[4, 5] One of Halsted's protégés, Harvey Cushing, published a case report in 1902 of a forequarter amputation of the upper arm for sarcoma with a pathological facture of the upper humerus. Under ether general anaesthesia, the brachial plexus was again directly visualised after surgical exposure and cocaine was applied to what would have been the supraclavicular brachial plexus. In a remarkable and thoughtful article, Cushing compares the rudimentary anaesthetic record ("ether chart" of pulse rate with annotations recording intra-operative events; reproduced in Figure 1.1) of his blocked patient, and argued that the term "regional anaesthesia" should be used to distinguish his technique from "local (infiltration) anaesthesia".[6]

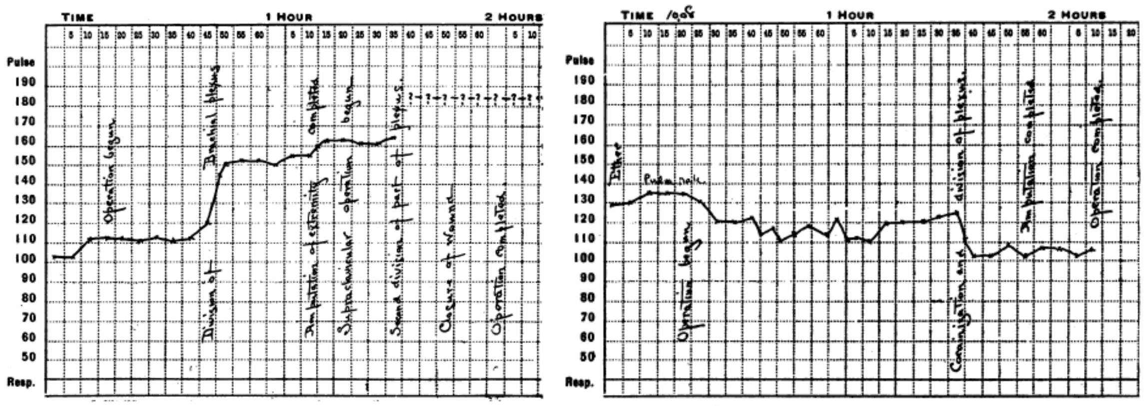

FIG. 4.—Chart recording pulse-rate during operation on Case I, December 26, 1899.

FIG. 9.—Ether chart recording pulse-rate kept during the operation on Case II.

Figure 1.1 Reproduction of ether charts from Cushing's article on "cocainization of large nerve-trunks". Intra-operative trends from one patient without a brachial plexus block, denoting the rise in heart rate with surgical excision (left), contrasting with stable heart rate if cocaine was pre-emptively applied to the plexus prior to upper limb amputation (right).[6]

By 1928, the German surgeon Diedrich Kulenkampff reported a case series of 1000 successful supraclavicular brachial plexus blocks performed by a percutaneous approach, illustrated in Figure 1.2.[7] He was methodical in describing his technique, choosing an appropriate needle, selecting 20 to 30ml of 2% procaine with a 1:200,000 adrenaline additive, and correct patient positioning. Kulenkampff carefully notes that the lung apex and subclavian vessels are adjacent to the plexus and "needless to say, a clear and definite mental picture of the anatomy of this region is absolutely necessary". The needle is inserted "over the midpoint of the clavicle, just external to the external jugular vein...the fingers press gently into the tissues over and close to the clavicle, and thus determine the pulsations of the subclavian artery". He advises his patients that they "must immediately notify the operator when the plexus is touched by the needle. [The patient] will recognise this by experiencing the same sensation as when striking his 'crazy' bone...with consequent tingling in the fingers". It is clear from this description that Kulenkampff relied on surface anatomical landmarks and palpation for the initial needle insertion point. Needle trajectory and endpoint prior to injection of local anaesthesia was ultimately guided by a paraesthesia technique that required direct needle-nerve contact.

FIG. 4.—Plexus anæsthesia in sitting posture. The direction of the entering needle is well shown.

Figure 1.2 Supraclavicular brachial plexus block by Kulenkampff. Patient positioning, needle entry point, needle trajectory, and ergonomics.[7]

Other techniques relied on tactile feedback from the advancing needle. This was dependent on subjective feel of a short-bevelled needle tenting into, and through fascial planes giving rise to loss of resistance, pops and clicks. These were limited to specific blocks (for example: epidurals, femoral nerve blocks, dorsal penile blocks), as the anatomy was conducive to have the needle pass through fascia or sheaths. Likewise, more definitive endpoints, such as flow of cerebrospinal fluid (intrathecal block) or aspiration of blood (transarterial axillary brachial plexus approach), were also constrained to specific blocks.

An alternative to the paraesthesia and tactile feedback techniques is motor nerve electrical stimulation using a peripheral nerve stimulator (PNS). The first clinical description of electrical stimulation of nerves occurred in 1912 by the German surgeon Georg Perthes, who used a nickel coated insulated needle to inject 2% procaine with adrenalin solution to provide analgesia in conjunction with ether anaesthesia during peripheral limb surgery. The needle was connected to an induction coil with the required electrical current modified by a foot operated pedal.[8] However, it was not until the 1960s that this technology was revived with the development of modern PNS equipment,[9, 10] but the technique was not routinely incorporated into RA practice until the 1980s.

Nerve stimulation is based on the principle that an externally applied, low current, low frequency, short duration electrical stimulus can depolarise Aα and Aβ motor nerve fibres. The rheobase and chronaxy of human motor nerves are well characterised, which in turn informs the typical settings of 0.5 to 1.5 mA current with pulse duration of 0.05 to 0.2 ms for peripheral nerve blocks, in conjunction with a fully insulated block needle.[11]

Opponents of the paraesthesia technique argue that safety is compromised due to the need for mechanical contact between needle and nerve,[12] the subjectivity of the patient's responses, and the discomfort it causes patients. They cite the appeal of peripheral nerve stimulation to include objective observation of motor twitch responses; electrophysiological principles of the current-distance relationship dictate that less current is required to depolarise a nerve, as the needle is brought closer to the nerve.[5, 13, 14] Aiming to elicit multiple nerve electrical stimulation versus multiple nerve paraesthesia technique has also improved rates of successful blockade to over 95%.[15]

Reality has shown that nerve stimulation is as much an art as a science, with some anaesthetists unable to replicate clinical results. Some of this inability to reproduce stimulation at specific thresholds were caused by the equipment: different needle designs have variable stimulating characteristics,[16] while PNS machines are inconsistent in delivering current, frequency and pulse duration settings.[17] Patient co-morbidities, especially peripheral neuropathies caused by diabetes, Charcot-Marie-Tooth, and chemotherapy, also increased electrical thresholds by an unpredictable level.[18, 19] Blood and fluid around the nerve can similarly increase the stimulation threshold.[13] Nerves themselves are heterogenous in composition of motor to sensory fibres, and the ratio of non-neural connective tissue to nerve axons increases with course from proximal to distal, with resultant changes in electrophysiological properties.[13]

These differences may help explain inconsistencies in motor stimulation thresholds found in a study by Choyce et al. Using paraesthesia as the primary nerve localisation technique, 77% of patients did have motor response at ≤ 0.5 mA but a significant proportion of the remainder required a current of 0.75 to 3.3 mA when stimulating nerves of the axillary brachial plexus.[20, 21] The inconsistency of stimulation thresholds were also observed in ultrasound studies involving the musculocutaenous nerve,[22] axillary brachial plexus,[23] and interscalene brachial plexus.[24, 25] PNS thus suffers from low sensitivity to needle-nerve proximity in clinical use.

The limitations and variability of nerve localisation using either paraesthesia or electrical stimulation restricted the widespread adoption of RA techniques. While RA was safe and reliable in experienced hands, the complexity of the technique meant that mastery of "regional anaesthesia will therefore continue to be the preserve of the committed few".[5]

1.2.2 The introduction of ultrasound-guided regional anaesthesia

The use of ultrasound for RA took discrete steps, beginning with identifying blood vessels known to be adjacent to nerves in the 1970s, then visualising the spread of local anaesthesia, before actual guidance of an inserted needle to identifiable nerve targets in the late 1990s. Clinically, all three methods remain viable options today.

In 1978, La Grange et al published their technique of performing the supraclavicular brachial plexus block with Doppler ultrasound.[26] Rather than relying on surface landmarks and palpation alone, a 9.5 MHz probe was used to echolocate the subclavian artery and vein, which was marked on the patient's skin (Figure 1.3). The probe was then removed before the needle was inserted and guided to the brachial plexus, with the endpoint confirmed by paraesthesia. This technique was thus properly termed an ultrasound-assisted block, where target nerves were not directly visualised by ultrasound, and the block needle trajectory was not guided by real-time ultrasound imaging.

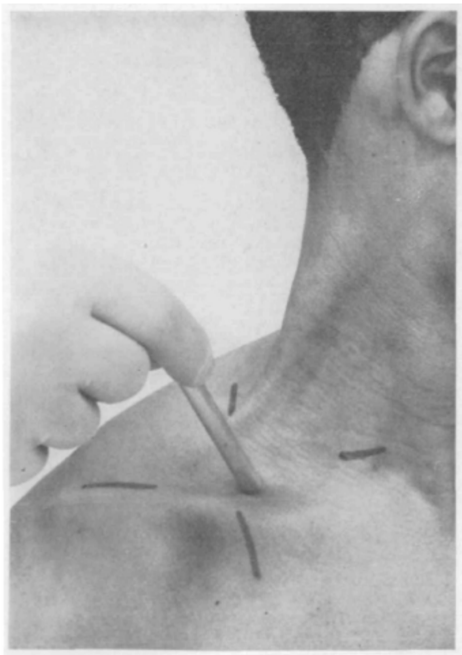

FIG. 1. Position of Doppler probe head and skin marks.

Figure 1.3 Ultrasound-assisted supraclavicular brachial plexus block by La Grange et al. The subclavian vessels are identified by Doppler ultrasound, before the completion of the block using paraesthesia to locate the brachial plexus.[26]

The next step required the technological breakthrough of B-Mode ultrasound. This allowed real-time 2-dimensional ultrasound images to be produced based on differential timing of reflected ultrasound echoes (representation of depth) and tissue interfaces (represented by intensity). B-Mode ultrasound was used by Ting et al in their 1989 paper to display the spread of local anaesthesia injectate around the axillary vessels after performing an axillary brachial plexus block.[27] While the authors performed the block using a surface landmark technique, their paper demonstrated some of the central features of UGRA: visualisation of vascular anatomy, and confirming appropriate hydrodissection of tissue planes by the injectate (Figure 1.4).

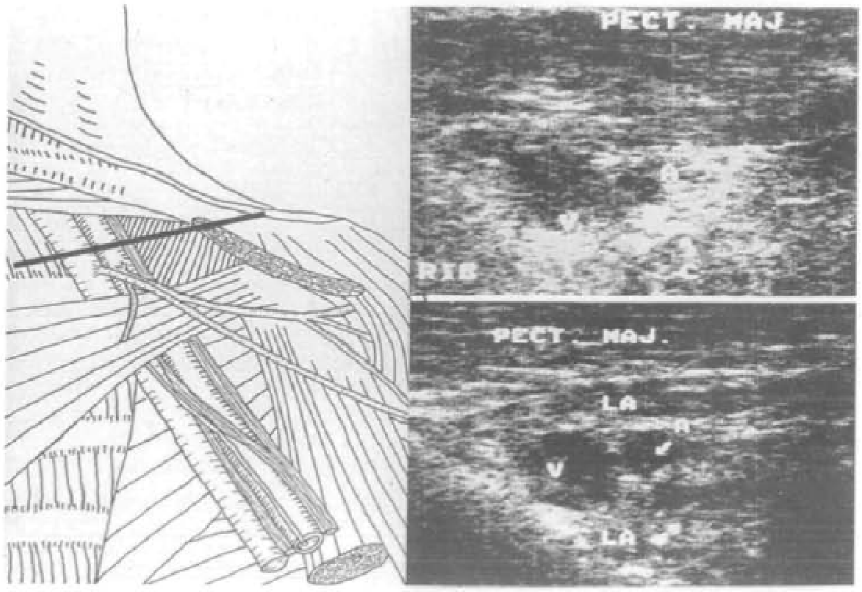

FIG. 2. Section across the apex of the axilla showing the artery (A) and vein (V) before (upper ultrasound picture) and after (lower ultrasound picture) the injection of 30 ml of 1.5 % lignocaine solution in water (LA). The artery and vein are better defined after LA injection, but are not separated. PECT.MAJ. = pectoralis major.

Figure 1.4 Spread of local anaesthesia (LA) around the axillary artery (A) and vein (V). Performed using a 3.5 MHz linear transducer. Nerves are visible but not recognised and not identified in Ting et al's paper.[27]

The first clinical description of an ultrasound-guided nerve block, where the target nerve structure was identified along with ultrasound visualisation of the needletip, was published by Kapral et al in 1994.[28] Hydrodissection of the tissue planes and injectate spread after supraclavicular and axillary plexus blocks was observed under ultrasound, and confirmed with radio-opaque dye. Subsequent publications in the first 10 years documented positive experiences using ultrasound guidance for interscalene brachial plexus blocks,[29, 30] infraclavicular brachial plexus blocks,[31, 32] and (the incorrectly termed) "3-in-1" blocks.[33]

These initial studies compared ultrasound-guided regional anaesthesia (UGRA) versus a traditional method, usually PNS or a surface landmark technique. Compared to the traditional method, UGRA demonstrated improved block quality and faster onset time. UGRA allows accurate placement of the needle adjacent to the target nerve, and the injectate can be directly visualised to confirm appropriate perineural spread.

Ultrasound-guidance has since spread beyond peripheral nerve blocks. Grau and colleagues published the initial study on the use of ultrasound to delineate the vertebral anatomy relevant for lumbar neuraxial blocks in 2001.[34] Previously, these blocks were solely performed with surface landmarks for initial needle placement, and needle trajectory was then informed by tactile feedback. Ultrasound has thus shed a window into this anatomical region, with further investigations on the sonoanatomy of the thoracic and lumbar spine[35] that improve the guidance and success of epidural, intrathecal,[36] and paravertebral blocks.[37]

In a similar fashion, novel blocks have been described that are only possible with ultrasound-guidance. Hebbard et al introduced the 'posterior approach' for the transversus abdominis plane block in 2007,[38] and the 'subcostal oblique approach' or transversalis fascia plane block in 2010.[39] Blanco published an approach to the pectoralis nerves for thoracic wall surgery in 2011,[40] and Visoiu and colleagues described a quadratus lumborum continuous catheter technique in 2013.[41]

In early 2016, the American Society of Regional Anesthesia and Pain Medicine (ASRA) updated their evidence-based review of clinical outcomes for UGRA blocks of the upper limb,[42, 43] lower limb,[44, 45] thoracoabdominal,[46, 47] and neuraxial[48] regions. These practice advisories concluded that ultrasound provides benefits of increased block success rates, faster onset of block, shorter block performance times, reduction in needle passes, and is at least non-inferior to other nerve localisation techniques in terms of risk of neurological complications.

Spurred by the emerging body of literature, two positive Cochrane reviews,[49, 50] and the above practice guidelines, UGRA has been quickly adopted as the primary modality for nerve localisation in regional anaesthesia, either alone or in combination with PNS as a secondary modality. One editorial has even claimed that any complications with UGRA performance is not a short-coming of the technology, but rather a deficiency of the proceduralist.[51] The technical and non-technical skills required of the proceduralist will be reviewed in the next section.

1.3 Procedural Skills in Ultrasound-Guided Regional Anaesthesia

Safe and successful performance of UGRA requires a combination of technical and non-technical skills. Most of the published literature for UGRA has focussed on technical skill sets. Some of these skills, such as ultrasound physics and "knobology" (understanding and fine tune adjustments of ultrasound machines to optimise performance) have common foundations with other medical craft groups utilising ultrasound.[52, 53]

Other skills include bi-manual dexterity required for successful transducer and needle manipulation under direct ultrasound visualisation. These skills have parallels with interventional radiology techniques, although the inclusion of the heel-toe transducer manoeuvre to better align the imaging plane with an in-plane needle approach comes from transthoracic echocardiography.[54] However, obtaining and optimising ultrasound images of nerves and musculoskeletal structures, and interpreting real-time spread of local anaesthesia injections, is particular to regional anaesthesia.

In 2010, ASRA and the European Society of Regional Anaesthesia and Pain Therapy (ESRA) published a joint statement and described the ten tasks believed to be necessary in UGRA procedures.[55] These tasks are listed in Table 1.1.

In comparison, non-technical skills have received minimal attention and until recently have not been explicitly acknowledged in training curricula. In a qualitative study, Smith et al found that experts in regional anaesthesia were not just technically proficient, but exhibited a capacity to adjust their schema and provide individualised management to patients.[56] This nuanced approach included contextually-appropriate levels of communication with patients and staff, anticipation of problems, and acknowledgment of one's own procedural limitations. Fletcher et al divides these non-technical skills into two further groups of cognitive/mental skills (decision making, situational awareness, resource management, planning) and social/interpersonal skills (communication, teamwork, leadership).[57]

In a recognition that this combination of technical and non-technical skills are desirable attributes of medical experts, the CanMEDS framework in Canada and later adopted by the Australian and New Zealand College of Anaesthetists (ANZCA), and the European Board of Anaesthesiology, Reanimation and Intensive Care,[58] have incorporated both these skill sets as learning objectives during postgraduate training.[59, 60] In the United States, a similar approach in developing a holistic curriculum was initiated by the Accreditation Council for Graduate Medical Education (ACGME) in 1999,[61] with consequent changes to anaesthesia resident training.[62] Table 1.2 is a summary of both technical and non-technical skill sets that are relevant for UGRA, drawn from primary source materials, the ASRA/ESRA recommendations, and ANZCA professional guidelines.

Table 1.1 ASRA/ESRA Recommendations for technical tasks in UGRA[55]

1	Visualise key landmark structures including blood vessels, muscles, fascia, and bone
2	Identify the nerves or plexus in short-axis imaging
3	Confirm normal anatomy and recognise anatomic variations
4	Plan for a needle approach that avoids unnecessary tissue trauma
5	Maintain an aseptic technique with respect to the ultrasound equipment
6	Follow the needle under real-time visualisation as it advances towards the target
7	Consider a secondary confirmation technique, such as nerve stimulation
8	When the needle tip is presumed to be in the correct position, inject a small volume of a test solution. If solution is not visualised during this test injection, presume that the needle tip is intravascular or out of the imaging plane
9	Make necessary needle adjustments if an undesired pattern of local anaesthesia spread is visualised. The visualisation of local anaesthesia should occur through the entirety of the injection to avoid an intravascular injection
10	Maintain traditional safety guidelines including the presence of resuscitation equipment, frequent aspiration, intravascular test dosing, standard monitoring, patient response, and assessment of injection characteristics

ASRA; American Society of Regional Anesthesia and Pain Medicine
ESRA; European Society of Regional Anaesthesia and Pain Therapy

Table 1.2 Summary of technical and non-technical skill sets for competency in UGRA

Skill set	Comments	References
Ultrasound Physics	Understanding principles of ultrasound imaging Recognition of ultrasound artefacts	Sites et al[52, 53]
Device operation	Selection of correct transducer Correct depth, focus, gain settings Colour Doppler function and application	Sites et al[52, 55] Marhofer[63] Brull et al[64]
Sonoanatomy	Identifying nerves, fascia, blood vessels Identifying bone, pleura and organs Lung sliding sign to exclude pneumothorax	Sites et al[53, 55, 65, 66] Litchenstein et al[67]
Image optimisation	Anisotropy Ultrasound frequency and depth penetration Short-axis versus long-axis views	Ihnatsenka et al[68] Chuan et al[69]
Motor skills – transducer movement	Pressure Heel-Toe Alignment Rotation Tilting	Sites et al[55] AIUM[54]
Motor skills – needle visualisation	In-plane needle technique Out-of-plane needle technique Hydrolocation Hydrodissection Bolus-Observe-Reposition Recognition of appropriate circumferential spread	Chin et al[70] Chuan et al[69]
Ergonomics	Appropriate location of self, equipment, patient	Sehmbi et al[71]
Equipment setup	Appropriate drug choice, needle, local anaesthesia Appropriate monitoring Infection control	ANZCA[72, 73]

Communication – patient interaction and consent	Open, effective, honest discussion of proposed block and material risks that a reasonable patient would attach significance	ANZCA[74]
Communication – Teamwork and using assistants	Block "Time Out" - right site, right patient Situational awareness and cultural appropriateness Workplace, occupational health and safety standards	ANZCA[60, 73, 75]
Post-block care	Medical record keeping Assessment and monitoring of patients in wards Care of continuous infusion catheters and pumps Appropriate transitional analgesia regime Appropriate hand-over to post-anaesthesia staff Follow up of side effects or complications	ANZCA[73]

AIUM; American Institute of Ultrasound in Medicine. ANZCA; Australian and New Zealand College of Anaesthetists.

1.4 Educational Theory and Acquisition of Medical Procedural Skills

There are many education theories that help explain the acquisition of cognitive and motor skills required for medical procedures. These can be divided into theories describing the learning patterns of individuals; theories describing the change in behaviours and motor skill proficiency as physicians progress from novice to competent performance; and theories describing specific educational interventions that improve procedural skill acquisition. A selected review of the more pertinent theories, and how they inform an evidence-based curriculum for ultrasound-guided regional anaesthesia training, will be discussed in this section.

1.4.1 Theories of individual learning patterns

A classic theory defining the requirements for learning is Maslow's hierarchy of needs.[76] As originally published in 1954, Maslow proposed 5 levels, composed of 4 basic needs (physiological, safety, love and belonging, esteem) and the fifth level of self-actualisation. Lower needs in the hierarchy need to be satisfied before there is motivation to progress to higher level needs. The original hierarchy has been modified with extra levels describing cognitive, aesthetic, and transcendence categories (Figure 1.5 and Table 1.3).

However, Maslow's theory has been criticised for being qualitative and based on interviews or biographies of narrowly selected individuals.[77] When measured across different human societies and cultures, Tay et al also found that Maslow's model is not strictly hierarchical as higher level independence and self-actualisation can be fulfilled independently despite lower social needs not being completely satisfied.[78]

Another theory of learning patterns is Kolb's 4-stage experiential learning style inventory.[79] The inventory can be viewed as both a classification system of learning, as well as describing the process governing learning. As a classification system, the four stages are denoted by the terms converging, diverging, accommodating, and assimilating, denoting the preferential style that an individual gains knowledge from experience (Figure 1.6 and Table 1.3). These specific styles are defined by two variables each taken from the perception continuum (a learner's preference to integrate new experiences through thinking, or via emotional responses to new experiences) and the processing continuum (a learner's preference to gain new skills through observation of the skill being performed, or via actively performing the skill themselves).

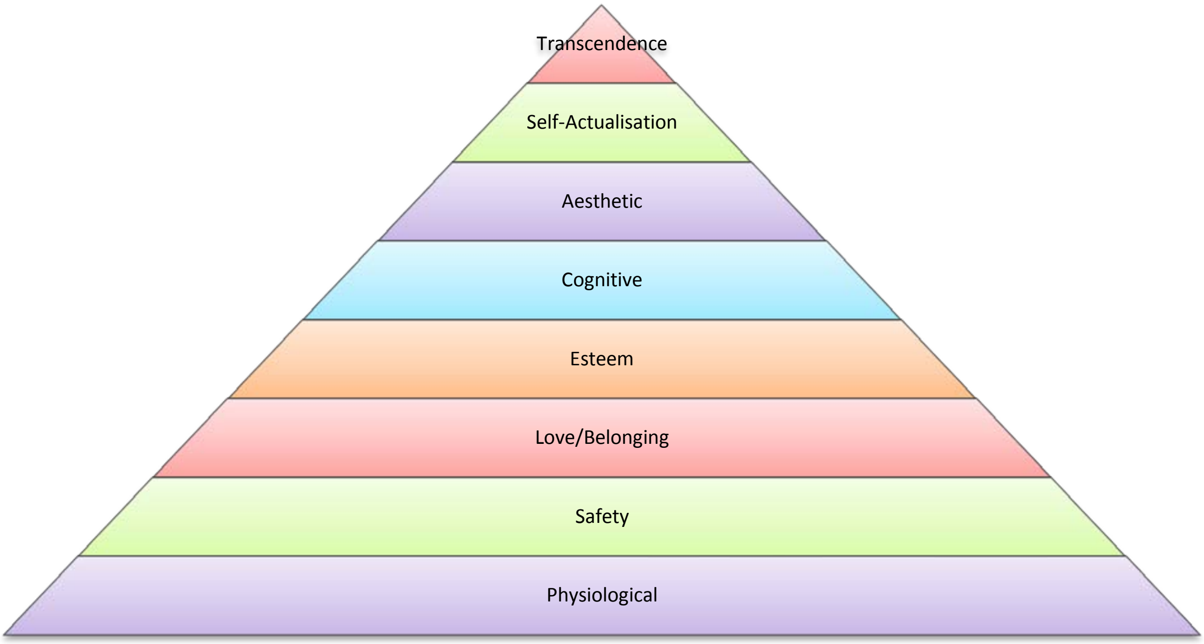

Figure 1.5 Maslow's hierarchy of needs. See Table 1.3 for description of levels.

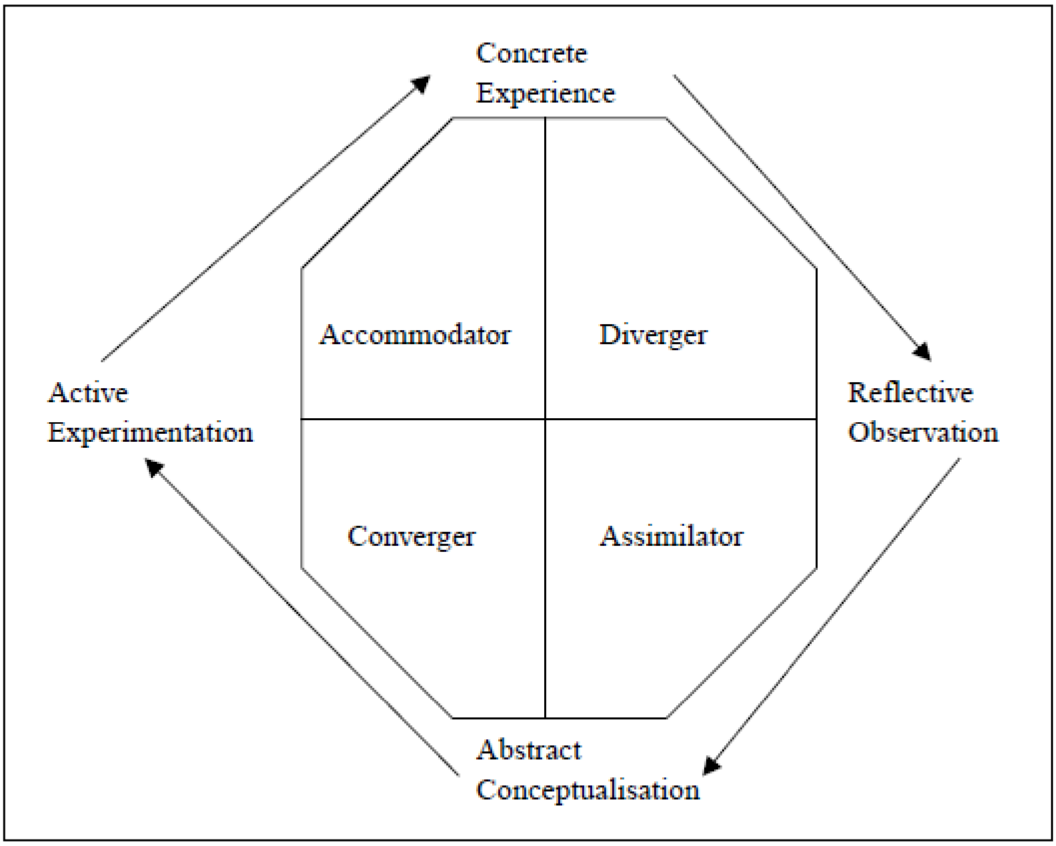

Figure 1.6 Kolb's experiential learning style inventory. Reproduced from Smits et al.[80]

Table 1.3 Theories of individual learning patterns

Theory	Description	Comment
Maslow's modified hierarchy of needs[76]	Physiological	Basic needs of food and shelter, sex
	Safety	Need for feeling safe, security
	Love and belonging	Friendship, intimacy, affection needs
	Esteem	Self-esteem, social status, independence and responsibility
	Cognitive	Desire for knowledge and meaning
	Aesthetic	Appreciation for beauty
	Self-actualisation	Realising potential, seeking personal growth, new experiences
	Transcendence	Desire to enable others to achieve self-actualisation, society-wide concerns
Kolb's experiential learning style inventory[79]	Converging style Abstract conceptualisation Active experimentation	Prefers to think and do Works best through practical workshops Likes problem solving
	Diverging style Concrete experiences Reflective observations	Prefers to feel and watch Works best in groups, listens to others Has multiple solutions to problem
	Assimilating style Abstract conceptualisation Reflective observations	Prefers to think and watch Works best with considered thinking Attracted to logical solutions
	Accommodating style Concrete experiences Active experimentation	Prefers to feel and do Works intuitively rather than analytically Likes new experiences and challenges

As a method to describe the process of learning, Kolb's inventory represents the four stages of a cycle that individuals must sequentially progress through to effectively learn new knowledge (Figure 1.6).[80] The four stages are anchored by the ends of the perception and processing continuums, namely concrete experience, reflective observation, abstract conceptualisation, and active experimentation. This theorises that a new skill is firstly experienced, causing the learner to reflect on that new experience, followed by analysis to integrate the new knowledge, which finally allows the learner to practically apply their new skill. Moreover, all four stages must be completed by learners for any new skill to be gained.

Lastly, trainers can draw upon specific cognitive development techniques initially described for childhood learning but has been applied to adult teaching. First described by Vygotsky[81] and elaborated further by Wood et al,[82] these complementary educational concepts include the zone of proximal development (ZPD), and scaffolding. ZPD describes the exposure of new skills to novices at a difficulty level that is commensurate with their level of experience. The task needs to be modulated to be appropriately challenging to activate their interest, but without being too complicated that would cause discouragement and frustration.[83] The matching concept of scaffolding is that trainers provide a level of support and guidance that enables the novice to meet the challenge at their ZPD. Practically, scaffolding can involve cues, prompts, hints, think-aloud modelling, controlling frustration, partial solutions, or direct instructions that allows the novice to gain insight into the next step of the procedure.[77]

This summary of individual learning patterns has relevance for UGRA training. Maslow's hierarchy suggests that trainers need to ensure a safe environment for learning. Apart from physical safety, it includes psychological safety whereby a trainee feels comfortable to ask for assistance. Delivery of UGRA teaching can be tailored by trainers, who judge the individual trainee's ZPD and provide an appropriate learning challenge, but offering effective scaffolding during the learning process. Over time, as the trainee gains confidence in the UGRA task, the scaffolding is withdrawn until the trainee is independent and competent. Understanding the learning style of the trainee is important to provide educational material that is best suited to their learning preferences. Ideally, training is conducted which provides opportunities for a trainee to process new skills through all four stages of the experiential learning cycle.

1.4.2 Theories on progression of motor skills learning

There are several theories on the behavioural attributes and motor performance characteristics of novices as they progress to becoming competent in procedural skills. While these theories describe specific stages of progression, motor skills are learnt in a non-linear fashion. Thus, theoretical stages should be considered as qualitative descriptions in a continuum of motor skills learning. Three of these theories are discussed below.

A well known theory is Miller's learning pyramid,[84] which postulates four stages of knows, knows how, shows how, and does, illustrated in Figure 1.7 and Table 1.4. The first two levels are theoretical and cognitive stages of learning, while the higher two levels show an increasing ability to translate theoretical knowledge into behaviours and motor skills that approximate competent performance.

There are several advantages when using Miller's pyramid to conceptualise motor learning. Firstly, each level is best examined using different assessment tools. Non-contextual multiple choice question (MCQ) exams are suitable for testing of recall of basic theoretical facts at the 'knows' level, such as the physics of ultrasound imaging. More clinically orientated MCQs should be used for the 'knows how' level, testing for example the ultrasound transducer movements, needle artefacts, and optimisation of sonoanatomy images. *Viva voce* examinations are suitable for both 'knows' and 'knows how' levels. The 'shows how' level requires initial demonstration of application of theoretical knowledge, and can be examined using Objective Structured Clinical Examinations (OSCE), such as a benchtop task to assess a trainee's real-time needle imaging skills on an *in vitro* phantom. The 'does' level is performance in an *in vivo* environment, and requires a workplace-based assessment (WBA) tool such as a clinically validated checklist for performance of a UGRA interscalene brachial plexus block in a patient undergoing shoulder surgery.

A second advantage of Miller's pyramid is assistance in designing a training curriculum. As higher levels are dependent on successful completion of the lower levels, the curriculum must address the requisite 'knowledge base' for each level. If objectives are clearly described for each level, this informs the goals of learning for the trainee and the suitability of the examination process.

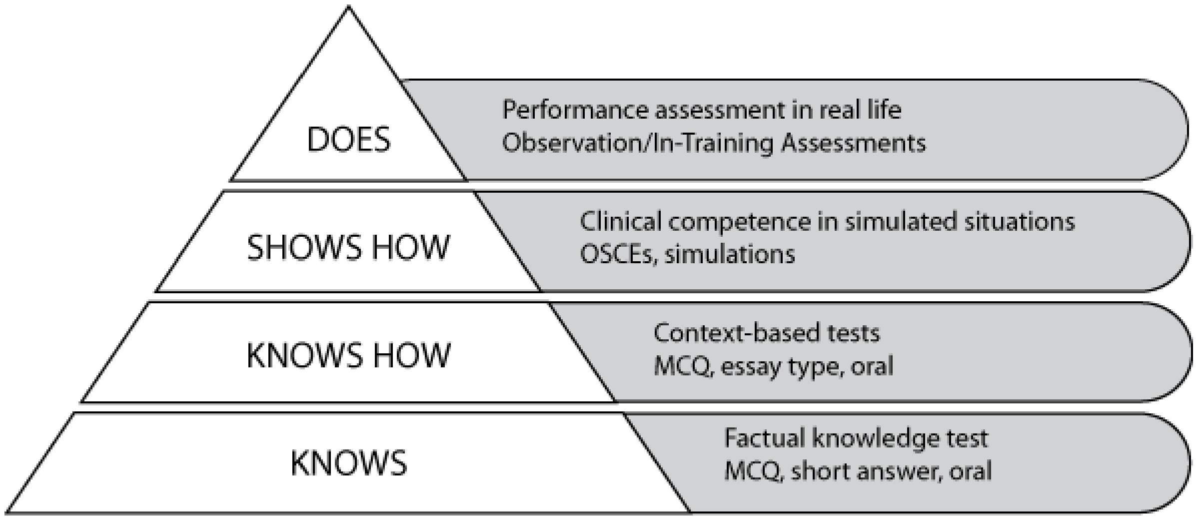

Figure 1.7 Miller's learning pyramid. Reproduced from the Australian Medical Council.[85]

OSCE; Objective structured clinical examination. MCQ; Multiple choice question

In 1967, Fitts and Posner published their multiphase theory on motor learning.[86] This theory describes three phases: a cognitive phase, transition to an associative phase, and lastly an autonomous phase (Figure 1.8 and Table 1.4). While the original theory was related to athletic skills, there is recognition of broad applicability to all motor skills including medical procedural performance. There is also supporting empirical and functional neuorimaging evidence for this theory.[83, 87]

Empirically, learning curves of motor skill performance versus practice follow a sigmoidal shape. Each of the three Fitts and Posner phases can be overlaid on the learning curve. The cognitive phase is characterised by a slow improvement in performance, with hesitant movements and numerous errors. With progression into the associative phase, performance is marked by rapid improvement towards the last autonomous phase. This is characterised by consistent high performance, with a plateau in motor skill improvement. Similarly, as motor skills are learnt different brain centres are observed to be activated, resulting in structural and functional changes when imaged with functional magnetic resonance imaging and positron emission tomography scanning.[88, 89]

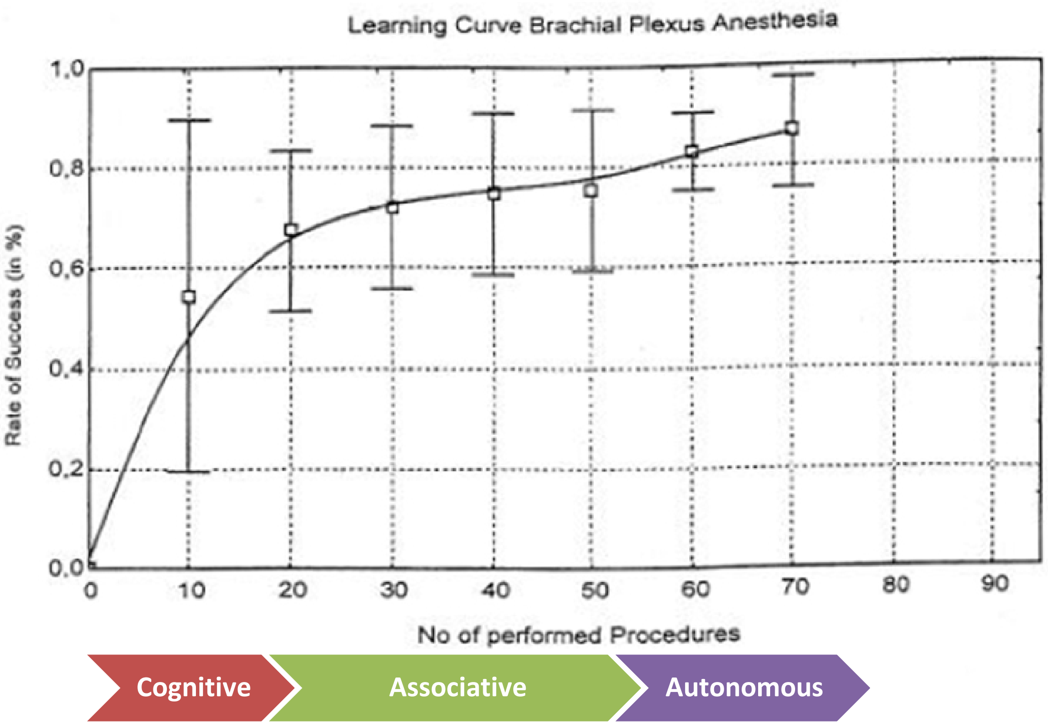

Figure 1.8 Fitts and Posner multiphase theory of motor skills acquisition. The three phases are overlaid on a learning curve derived from observations of novice anaesthesiology residents performing axillary brachial plexus blocks.[90]

The third theory to be reviewed is the Dreyfus and Dreyfus 5-stage model of skill acquisition published in 1980, derived from observations on expert chess players and air force pilots.[91] The original model has since been modified, and in this version the five stages are novice, advanced beginner, competency, proficiency, and expert. This is illustrated in Figure 1.9 and Table 1.4.[92]

The Dreyfus model was used as the framework for the 2009 physician curriculum published by the American Board of Internal Medicine and ACGME.[93] For anaesthesia trainees, the latest 2015 Milestone project between the American Board of Anesthesiology (ABA) and the ACGME explicitly referenced the competency stage (Level 4) as the minimum standard for graduation and transition into independent practice.[94] The milestones, or learning objectives, for regional anaesthesia is shown in Table 1.5.

Table 1.4 Theories of motor skill progression

Theory	Description	Comment
Miller's learning pyramid[84, 95]	1. Knows: Gathering facts	Theoretical knowledge
	2. Knows How: Interprets theory	Application of knowledge
	3. Shows How: Demonstrates learning	Demonstration of performance
	4. Does: Performance in workplace	Independent clinical practice
Fitts and Posner multiphase theory[83, 86]	1. Cognitive Learning new motor skill Step-by-step motor performance	Movements are error-prone, inefficient, indecisive Lack of situational awareness
	2. Associative Development of a schema Inefficient movements rejected Different methods trialled Practice and feedback required	More fluid, efficient movements Some automaticity of motor skill Less attention required for motor tasks, allowing improved non-technical skill performance
	3. Autonomous Expert skill level	Accurate, consistent, smooth Little cognitive attention required Adapts easily to unexpected events
Dreyfus and Dreyfus 5-stage model of motor skill acquisition[92, 96]	1. Novice Rules/textbook based performance	Performs skill without context Rigid application of rules
	2. Advanced beginner More flexibility in rules, autonomy	Can complete straightforward tasks Understands complexity but not resolve
	3. Competent Autonomous, takes responsibility	Organised decision making Good working knowledge and skill base
	4. Proficient Deeper understanding	Intuitive, routine high performance Recognises patterns by experience
	5. Expert Authoritative understanding	Able to create new paradigms of skill Holistic and analytic problem solving

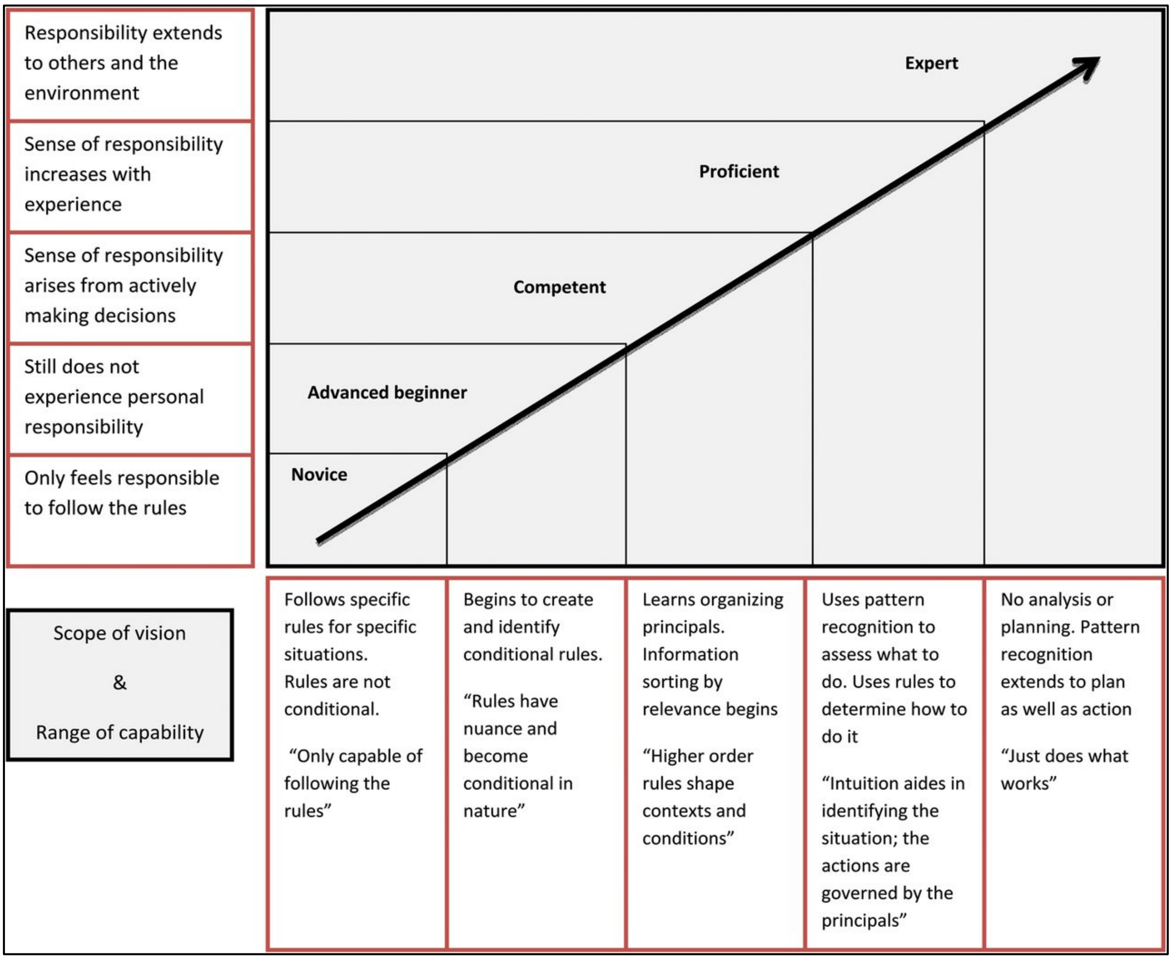

Figure 1.9 Dreyfus modified 5-stage model of skill acquisition. Reproduced from Kirkpatrick et al.[97]

Thus, the aim of postgraduate training is to produce physicians with the minimum following characteristics: an ability to modify rigid rules and prioritise according to context and for individual patients; while clinical decision making is not as refined or as intuitive as more expert proceduralists, procedures are safely performed and the practitioner can undertake most tasks autonomously without supervision. Importantly, when faced with more demanding clinical scenarios, there is a hierarchical and logical organisation for problem solving. There is a good working background of knowledge, though not a depth of understanding that is the feature of expert performance. Nonetheless, the physician has gained sufficient clinical exposure and professional acumen to continue learning post-graduation to higher levels of proficiency.

Table 1.5 Milestones for regional anaesthesia competency. Reproduced from the 2015 Anesthesiology Milestone Project of the ABA and ACGME.[94]

Level (Year of training)	**Milestone description**
Level 1 First post-graduate year; Prior to entering CA1 year of anaesthesiology	Demonstrates sterile technique Administers infiltrative local anaesthetics for procedures under direct supervision Identifies physiologic changes associated with local anaesthesia administration and seeks help appropriately
Level 2 Resident in anaesthesiology; Prior to subspecialty training	Applies appropriate monitors and prepares resuscitative equipment prior to performing regional anaesthesia procedures Performs spinal and epidural anaesthesia under direct supervision Recognises problems or complications associated with regional anaesthesia, and manages them under direct supervision
Level 3 Has experience in subspecialty training	Performs peripheral nerve blocks and regional anaesthesia under direct supervision, including both upper and lower limb extremity blocks and thoracic epidurals Uses ultrasound or nerve stimulator guided techniques appropriately Performs common paediatric regional anaesthetics (eg. caudal blockade) with direct supervision Recognises problems or complications associated with regional anaesthesia and manages them with indirect supervision
Level 4 Fulfils expectations of independent practice Graduation target	Performs spinal, epidural, and peripheral nerve blocks with conditional independence Supervises junior residents in performing regional anaesthetics and other health care providers on issues related to regional anaesthesia Manages problems or complications associated with regional anaesthesia with conditional independence
Level 5 "Aspirational goals", beyond performance targets expected for residency	Independently performs peripheral and neuraxial regional anaesthesia techniques Independently manages problems or complications associated with regional anaesthesia

1.4.3 Teaching strategies for motor skills acquisition

Specific educational strategies have been postulated to improve the quality of motor skills training. In this section, the following strategies will be examined: cognitive task deconstruction/analysis and fractionation, *in vitro* pre-training, deliberate practice versus discovery learning, and structured feedback.

In task deconstruction, a systemic approach is taken to decompose a complex procedure into component tasks.[98] This has several advantages: it facilitates teaching of novices by providing simpler tasks to initially learn and practice; and each task can be assessed independently instead of the complete procedure. This process is called fractionation, while the re-combining of separately learnt motor skills back into the complete procedure is called segmentation.[83] An example in UGRA would be learning the sonoanatomy task (transducer movements, interpretation and recognition) independently from the real-time needle guidance task, before attempting to perform a clinical nerve block procedure.

Several methods can be used to perform task deconstruction. The simplest method is expert opinion of tasks considered necessary in a procedure. However, unintentional omission of 70% of critical tasks can occur.[99] Experts' skills and knowledge have become largely autonomous, and they often employ short-cuts in performance that cannot be replicated by novices.[98]

To avoid this bias, researchers have used cognitive task analysis. This employs semi-structured interviews and cross-referencing against observations of experts to compile a "gold-standard" decomposition of procedures. This form of analysis has created deconstructed tasks for femoral artery Argyle shunting,[99] endoscopic retrograde cholangiopancreatography,[100] central venous cannulation,[101] and colonoscopy procedures.[102] O'Sullivan et al used hierarchical analysis to deconstruct an ultrasound-guided axillary plexus block into 256 smaller tasks.[103] However, application of the identified tasks to a teaching curriculum is limited by the poor inter-rater agreement (Cohen's kappa < 0.1) between the opinions of the two experts (anaesthetist and clinical psychologist) used in this study.

After task deconstruction, novices are trained in the individual fractionated motor skills. Ethically, initial practice should use *in vitro* or benchtop models rather than on patients, as novice errors are very common in the early learning phase. Ideally, once skills have been objectively examined as safe and competent, novices are then permitted to perform clinical procedures. This concept is called pre-training, and takes into account that basic psychomotor and cognitive skills have now been trained to a level of automation. The novice has now more attentional resources to effectively learn additional skills required for procedural performance.[104]

Deliberate practice is a concept introduced by Ericsson when defining the characteristics of expert performance. Through research on professional musicians, chess players, and across multiple athletic sports, expertise was gained through intentional and purposeful practice on well-defined tasks with objective goals, immediate and detailed feedback from trainers that can be used to improve elements of performance, and repetition of the same or similar tasks.[105]

Deliberate practice is therefore a focused type of learning that incorporates provision of effective feedback, practice by motivated learners who strive to achieve pre-defined learning outcomes, repetitive practice, and practice with different levels of difficulty.[106-108] In a meta-analysis of trials which incorporated deliberate practice as the educational intervention, there were significant improvements in novice performance for central venous catheter insertions, laparoscopic skills, advanced cardiac life support, cardiology examination skills, and performing thoracocentesis.[109] In a randomised controlled trial, novice surgeons were trained in laparoscopic cholecystectomy in a simulation laboratory using conventional teaching or deliberate practice techniques. When measured using three separate validated objective structured assessment of technical skills (OSATS) global rating scales, novices trained using deliberate practice had significantly more proficient performance.[106] This replicates results from an earlier study using a cadaveric porcine model,[110] but also in airway management,[111] and weaning off cardiopulmonary bypass.[112]

This is compared with discovery learning, which postulates that skills are best learnt when novices draws on their own experiences and prior knowledge. Emphasis is placed on experimentation rather than direct external instruction. In a study examining the learning curve of novices being taught real-time ultrasound-guided needle manipulation skills, de Oliveira Filho et al extrapolated that 109 attempts were required for 95% of novices to successfully learn how to direct a needle to a target structure under ultrasound.[113] Novices were provided with basic information on ultrasound skills but no other feedback or instructions were given by supervisors, as part of a discovery learning technique. In contrast, with the same learning objective and similar statistical methodology (cumulative sum with Bush and Mosteller's mathematical modelling), Barrington et al found that novices with deliberate training were able to achieve the same outcome with only 28 attempts.[114]

An important feature of deliberate practice is structured feedback. Slater et al[83] modified a schema originally proposed by Nicol and Macfarlane-Dick,[115] and describes an internal and external pathway for trainers to provide feedback to trainees learning new motor skills.

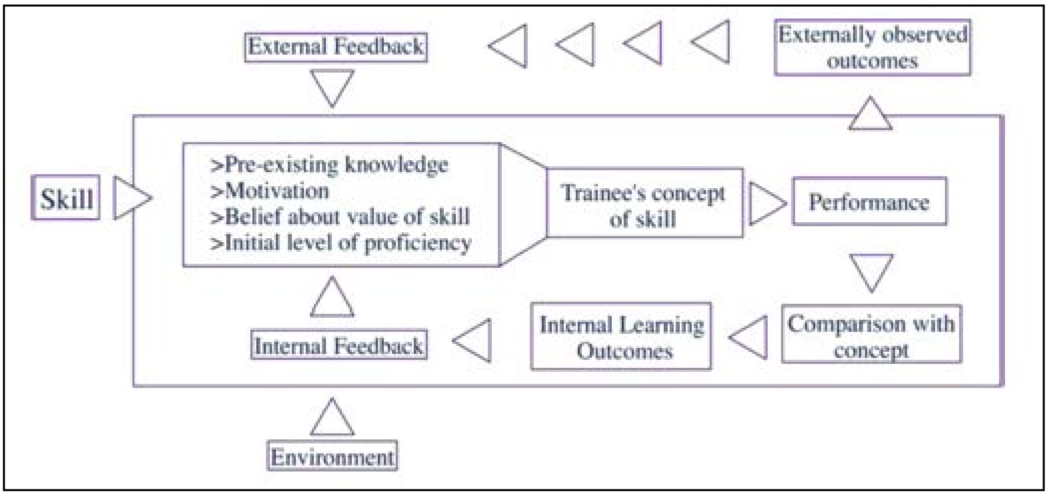

Figure 1.10. Internal and external feedback pathways in deliberate practice. Reproduced from Slater et al.[83] The trainer also conducts pre-briefing and debriefing before and after the performance as part of structured feedback.

Actual performance is compared internally by the trainee against their concepts of how the skill should have been performed. This can be validated with, or contrasted against, the external feedback from the trainer (Figure 1.10).

In addition, trainers can conduct a briefing and debriefing discussion with the trainee, which sets out the learning objectives, expectations, and responsibilities during the procedure. This type of structure is a prompt to give permission for external feedback to occur, and is particularly relevant for clinical environments. Without creating a deliberate and protected opportunity to teach, distractions due to production pressure and clinical considerations reduces the educational impact from learning in the operating theatre.[116]

These individual educational strategies have been incorporated into a generic framework for systemic training and assessment of technical skills (STATS).[98] Learning objectives are identified for both theoretical and practical skills. Theoretical skills can be examined using appropriate assessment tools, typically written MCQs, short answer, and/or oral examinations. Task deconstruction of a complex motor skill helps define the component steps that are considered essential for the procedure. Once identified, these fractionated steps can be practiced separately in the form of part-task *in vitro* pre-training. While not representative of the whole, part tasks allows training in a smaller subset of theoretical and motor skills that reduces the risk of cognitive overload. Once confidence in smaller tasks has been achieved, multiple tasks can be assembled into performance of more complex procedures.

Trainers can employ principles of deliberate practice to reduce the time taken to reach competency. This entails the use of structured feedback, briefing and debriefing between trainer and trainee. These skills should then be transferred to a clinical environment. In both the *in vitro* and clinical environments, evaluation of competency requires validated and reliable assessment tools. Importantly, volume of practice or logbook metrics should not be used as basis for competency.[98] Satisfactory assessment scores at the end of training allows the granting of clinical privileges. This is shown in Figure 1.11.

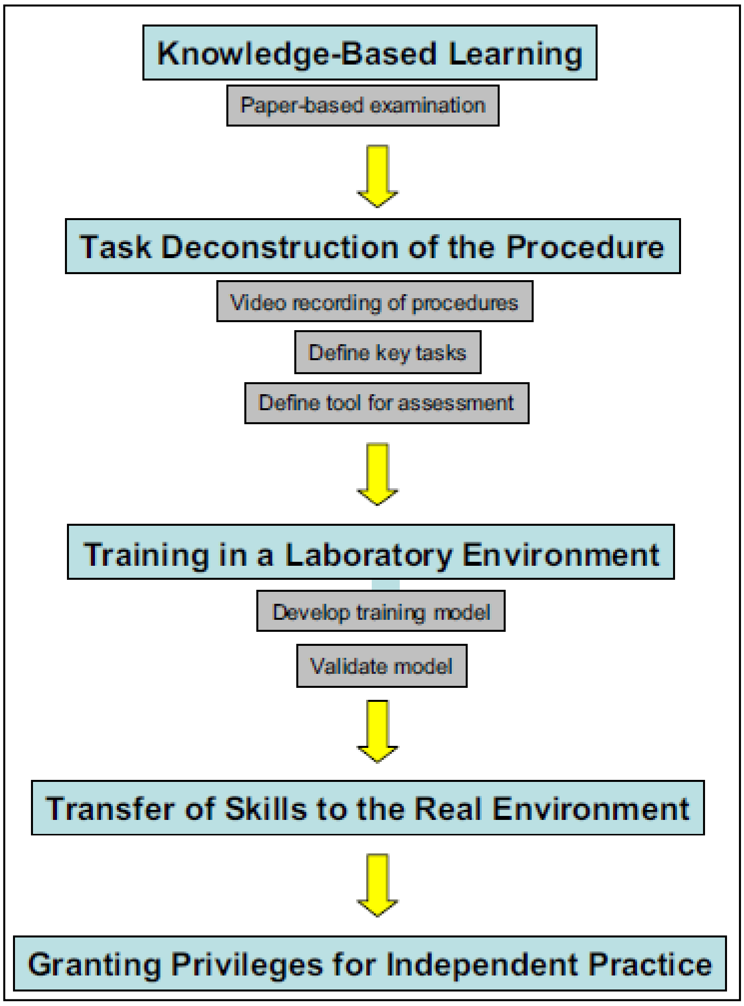

Figure 1.11. A framework for Systemic Training and Assessment of Technical Skills (STATS). Reproduced from Aggarwal et al.[98]

A practical expression of the STATS framework has been published for the redesigned regional anaesthesia curriculum at the Mayo Clinic College of Medicine.[117] The important features of their curriculum, and parallels with the STATS ideals, are summarised in Table 1.5. The types of assessments, and level of clinical supervision given to their anaesthesiology residents, was unclear in this paper. For example, whether feedback was continuously provided during the clinical placement was not described. The format of assessment was also not specified for both pre-training and clinical phases. Nonetheless, this commendable redesign incorporates many of the educational concepts with an evidence base to improve UGRA education and training.

Table 1.5 The 2009 Mayo Clinic regional anaesthesia curriculum[117]

Curriculum Feature	Description	Comments
1. Standardised educational content	All residents receive a curriculum pack prior to commencing rotation: - Pre-anaesthesia assessment and pre-operative workup - Surgical and Block site time-out - Sterility and infection control - Regional anaesthesia technique - Appropriate monitoring - Intra-operative care - Post-operative care - Medical note documentation	Knowledge based learning Learning objectives articulated for each task Assessment of knowledge not specified
2. Suggested reading and resources list	Internet based learning modules Multimedia videos	Online assessment of theoretical knowledge
3. Pre-clinical training	*In vitro* training program: - Session 1: Ultrasound physics and equipment - Session 2: Scanning techniques and sonoanatomy - Session 3: Sonographic needle guidance - Session 4: Demonstration of sonographic proficiency	Task deconstruction Each session has defined learning objectives Feedback not specified Session 4 is assessed using a checklist; not further specified

	Simulation laboratory training: Management of complications, and training in non-technical skills	Examples include management of LAST. Feedback not specified
	Anatomy laboratory training: Cadaveric dissection	Experiential learning
	Online simulation training: Virtual regional anaesthesia performance	Ability to practice techniques
4. Clinical Assessment	Multisource feedback on learning outcomes: - knowledge base - clinical judgment - technical skills - patient rapport - professional interaction with staff - organisational skills	Monthly discussion of each resident's performance at department level Supervision on daily level not specified Feedback not specified
	Residents provide feedback on faculty	Quality assurance for training

LAST; local anaesthesia systemic toxicity

1.5 Assessing Medical Procedural Skills

The process of examining and granting clinical privileges to medical trainees has evolved greatly in the last half century. Previously, assessments were based on *viva voce* examinations and written essays, with the emphasis on knowledge content and factual recall.[118] Assessment of procedural skill was through a combination of volume of practice (logbook) cases, audits (complication rates, mortality/morbidity data), and supervisor appraisal. Supervisor assessment was typically retrospective, occurred once at the end of rotation, and did not use defined criteria.

The fundamental change in postgraduate anaesthesia education has been moving from this traditional apprenticeship model to a competency-based model, as articulated by the current ANZCA, United Kingdom Royal College of Anaesthetists (RCA), European Board of Anaesthesiology, and ABA/ACGME curricula. Competency is also contextual; basic principles and core concepts for anaesthesia are introduced early in training and expanded continuously throughout training to keep with the increasing knowledge, skills and maturing behaviours of the trainee. This is termed 'spiral learning' and is explicitly referenced in the ANZCA and RCA curricula.[119, 120] The traditional and competency-based education models are compared in Table 1.6. These curricula changes have implications for the nature, type, and location of assessment; and requires creating an environment to empower trainees to be active participants in their own training.

The nature of assessment has changed to expand beyond the original purpose as purely barrier entrance and exit exams, to include assessments that allow provision of deliberate feedback, based on educational concepts outlined in the previous section. These two different roles are respectively called summative and formative assessments, which have been described as "assessment *of* learning" and "assessment *for* learning".[121] Several types of assessment have also been introduced, with different tools created for specific summative and formative purposes. Thirdly, trainees are being assessed outside the traditional examination hall, with a desire to measure performance in the simulation laboratory as well as in clinical environments such as hospital wards and operating theatres.

Another significant change is the primacy of feedback. Effective feedback is integral to deliberate practice as described in the previous section, and is the basis of formative assessment. This section will discuss summative versus formative assessment, and the evidence for the educational value of providing feedback. The types of assessment tools, and the process of validating them for clinical use, will be reviewed in later sections.

Table 1.6 Comparison of apprenticeship versus competency-based models of postgraduate medical education[59, 61, 122]

Apprenticeship model	Competency-based model
Relationship structure	
Hierarchical; consultant-trainee	Facilitator; coach-learner
Assessment role	
Summative Assessment of learning Norm based (relative to previous trainees) Personal opinion Acquired knowledge, skills based on exposure	Formative and summative Assessment for learning Criterion based (defined attributes/objectives) Objective measurement Demonstrates knowledge, skills against outcome
Assessment type	
Essays and short answer written exams Viva voce examinations Rarely assess non-technical skills	Also includes purpose designed tools: Checklists, GRS, DOPS, OSCE, OSATS Includes assessment of non-technical skills
Assessment location	
Examination hall	Also includes simulation laboratory Workplace: wards, theatres, clinics
Timing of assessment	
Retrospective End of rotation and single occasion High stakes barrier each rotation	Direct observation Prospective and regular Portfolio of assessments over entire training
Feedback	
Unstructured	Briefing/debriefing Deliberate Structured

GRS; global rating scale, DOPS; Direct Observation of Procedural Skills

OSCE; objective structured clinical examination,

OSATS; objective structured assessment of technical skills

1.5.1 Summative and formative assessment

In 2010, the Ottawa Conference on the Assessment of Competence in Medicine and Healthcare Professions published consensus statements on definitions and criteria for assessments.[123] Summative assessments are high stakes examinations, with the primary rationale to satisfy requirements of credentialling and fitness to practice by ensuring all graduating anaesthetists have met a standard of knowledge, safety, and professionalism. The credibility of summative assessments rests on the test producing reproducible, valid, and fair results. Less consideration is given to logistical feasibility and educational value of the test itself.

By comparison, the primary purpose of formative assessments is a stimulus that "provides results and feedback in a fashion that create, enhances, and supports education; it drives future learning forward".[123] Formative assessments must be valid, highly feasible to use within the competing demands of a busy workplace, and creates opportunities for teaching, while still retaining some reproducibility.

Impediments to successful use of formative assessments includes scheduling of assessment alongside the pressures of clinical workloads, and lack of training of assessors.[121] While a one-day program for anaesthetists to be trained as assessors was unsuccessful in improving their reliability of scoring,[124] more intensive training did allow trainers to score more consistently using the mini-clinical evaluation exercise (mini-CEX) assessment tool.[125]

Defining the purpose of assessment therefore determines which properties of an assessment tool should be emphasised, and a compromise is required between competing intentions.[126] Balancing the mix and timing of summative and formative assessment remains challenging for designers of a curriculum.[127]

In cognitive psychology, the 'testing effect' describes the behaviour that the act of assessment forces trainees to actively reinforce memory and cognitive organisation ("effortful processing"), thus promoting learning.[128] This is the basis for the axiom that "assessment drives learning". While previous literature have focussed on the quite different purposes of summative and formative assessments, the testing effect between the two types of assessment has only recently been studied.

In a quantitative analysis of 25 out of 27 United Kingdom medical schools, Devine et al measured the number of hours and intensity of summative assessment at undergraduate level and compared to academic success at postgraduate specialty training. They found a strong and significant correlation between more hours spent assessing medical students and their scores at knowledge and clinical examinations at postgraduate examinations.[129]

In two randomised controlled trials, from the same group of researchers, summative assessments were compared against formative assessments on their impact on electrocardiography interpretation skills. Summative assessments were stronger drivers than formative assessments in improving retention in the short-term[130] and medium-term, while intensity of training was not significant.[131]

In another randomised controlled trial, medical students were taught advanced cardiac life support skills. Both control and intervention groups received formative assessment, but the intervention group also received a summative assessment using a validated checklist anchored on European Resuscitation Council guidelines. This intervention group performed significantly better at the end of training versus those who only received formative assessments.[132]

These are perhaps intuitive results (higher stakes examinations driving better retention of knowledge and skills) but no studies have been performed in anaesthesia, and their significance for the mix of summative and formative assessments in a curriculum remains to be determined.

1.5.2 Feedback

Feedback is a process by which a procedural performance is analysed, questioned, reflected, and reframed in order to learn or improve performance.[77] Practically, the feedback can inform trainees of their level of progress, their learning weaknesses and direct them to potential resources to assist in remediation, and motivates trainees to learn.[133] As discussed in the previous section, feedback can be internally generated from the trainee or provided by an external trainer. As part of deliberate practice, feedback constitutes a core component of gaining expertise.

In a systematic review on the importance of feedback on learning, the majority of studies found feedback was a positive educational strategy, either when used as the sole intervention, or when combined with other educational interventions.[134] The source and repetition of external feedback was significant, with more positive effects seen with feedback given by authoritative sources over an extended period of time.

In another systemic review, high versus low fidelity simulation was compared for their effects on teaching technical skills. Instead, the authors found that feedback was more important than fidelity of the simulator.[135] Feedback can take place in the context of deliberate practice, or be automatically provided by the simulator, typically by comparing the novice's performance against an expert's performance. Interestingly the second most effective learning strategy was task repetition, again underscoring the principle of deliberate practice. In contrast, the fidelity of the simulator is rated low as a contributor to effective learning.

Another advantage of deliberate practice with feedback was longer skills retention after training. Using a global rating scale and hand motion analysis with the Imperial College Surgical Assessment Device, novices who received verbal feedback after each attempt showed less decay of newly learnt basic surgery skills than those trained with discovery learning alone, or discovery learning with hand motion analysis comparison to expert performance.[136]

The timing of feedback has received some attention. Concurrent feedback is delivered during the procedure, while summary feedback is provided at the end of the procedure. Concurrent feedback is intuitively more helpful for novices who are processing information in the Fitts and Posner cognitive phase or Dreyfus novice stage.[83] However, in a study using novice medical students, summary feedback resulted in better retention of basic surgical skills than concurrent feedback.[137]

Despite the importance of feedback, Norcini and Burch note that trainers often fail to provide effective feedback. One reason included the precedence of numerical scoring of workplace-based assessment tools over giving feedback; this could be due to curriculum issues (role of assessment was not clarified as summative or formative), structural issues (there was insufficient space in the WBA form to provide feedback), and trainer issues (inadequately trained to understand the importance of feedback).[133] The second reason is lack of buy-in by the trainers; without having ownership or regular engagement in the educational process, it is unlikely that effective feedback will occur. Lastly, there is evidence that there is a wide variability in the quality of feedback given by trainers. Some of the factors leading to this variation is only beginning to be understood, and includes a complex interaction between personality styles, method of querying, perceived roles, emotional state, experience, training, and confidence in their ability to provide feedback.[138-140]

1.6 Psychometric Properties of Assessment Tools

In the preceding section, summative and formative assessment tools are explained as essential components of current anaesthesia training. The design, structure, and quantitative characteristics such as reliability and validity, are called the psychometric properties of an assessment tool.

Evaluating the psychometric properties of each tool ("assessing the assessment tool") is necessary for acceptance by all stakeholders. This includes confidence in the results of the tool by both trainees and assessors; how accurately the tool measures trainees against the competency definitions and skill sets described in the curriculum; is the assessment robust to tolerate external scrutiny by professional bodies and societal standards; are the results reproducible; and is the assessment objective. Validated tools are also needed in education research to measure performance changes after training interventions, and for serial measurement of trainees over time.[141]

Ideally, before tools are included in curriculum designs, pilot testing has occurred and their psychometric properties defined.[95] However, ongoing testing of the tool is still required after implementation to provide quality assurance of stable validity and reliability characteristics. This is summarised by the ANZCA Guidelines on Assessment that "relevant psychometric data on existing and new assessments should be routinely collected and acted upon".[142] These include testing of validity and reliability over time with different trainee populations and test conditions.[143] Due to these limitations, results of validation studies done on specific participant populations and procedural tasks should not automatically be generalised to all trainees in the curriculum or to other anaesthesia procedures.[144]

In this section, the psychometric properties of an ideal assessment tool is discussed, as well as the statistical methods to quantify reliability.

1.6.1 Properties of an ideal assessment tool

Van der Vlueten first proposed an utility index of assessment tools in 1996,[145] with subsequent endorsement and elaboration in the 2010 Ottawa consensus statement on assessment tools.[123] These descriptions are summarised in Table 1.7 as properties of an ideal assessment tool for UGRA.

While these attributes are consistent with an ideal tool, practical design and application of assessment tools require compromises. In UGRA, a wide range of knowledge, technical and non-technical skill sets are required. It is unrealistic for a single assessment tool to comprehensively evaluate a trainee's complete set of UGRA knowledge and skills.[141] Every assessment tool is therefore is a balance between competing demands. By first identifying the intention of assessment, this helps inform which psychometric property is emphasised during the design phase. Equally, clinical use of existing assessments should be guided by the original design intent of the tool. While a summative assessment tool (high stakes examination of knowledge or skill) could be repurposed for formative assessment (providing a stimulus for feedback), it may be difficult for the reverse to occur given the former's requirement for high reliability.

The interdependent nature of psychometric properties is most pertinent for validity and reliability. Validity has been argued as "an ongoing process of hypothesis generation, data collection and testing, critical evaluation and logical inference".[146] Reliability is defined as reproducibility of test scores, especially in regard to consistency of scoring between different assessors who view the same trainee performing the same procedure. The terms between-assessor, inter-rater agreement, and external reliability are equivalent.

The minimum level of validity acceptable for a competency-based UGRA assessment tool requires evidence that the skill sets of that procedure is examined (content validity), and the tool can differentiate trainees who are performing the procedure satisfactorily (concurrent validity). In turn, justification of validity requires quantitative statistical analysis of how consistently test scores are obtained,[144] and it has been noted that "reliability is a major source of validity evidence for all assessments".[147]

Table 1.7. Psychometric properties of the ideal assessment tool for UGRA.

Adapted from Van der Vlueten,[145] Norcini et al,[123] Bould et al,[121] Ahmed et al,[148] Gallagher et al.[149]

Property	Description and comments
Validity	Tool measures what it purports to measure Evidence for validity is built up over multiple evaluations and in different contexts. Body of evidence then supports validity. Must demonstrate reliability as a core feature of an UGRA assessment tool
Face validity	Suitability of tool to the real-life UGRA clinical reality Determined by subjective expert opinion
Content validity	Relevant UGRA knowledge or skill set is being tested Determined by subjective expert opinion
Construct validity	Test scores differentiates inexperienced and experienced proceduralists
Discriminate validity	Test scores can differentiate levels of novice versus expert ability within the same cohort of UGRA proceduralists
Concurrent validity	Scores compared with a gold standard tool; however no such tool exists. As an alternative standard, test scores are consistent with external and objective measures of block performance, such as success, time to onset, number of needle passes, complication rates
Predictive validity	Test scores are predictive of actual clinical outcome. For example, test scores from an *in vitro* or simulation environment are predictive of future or actual clinical UGRA performance
Reliability	Reproducibility of test scores if repeated under similar circumstances
Internal reliability	Different trainees scoring consistently in the same section of the tool Not as important as external reliability Tests used to measure internal reliability have included: Split-half reliability Cronbach's alpha
External reliability	Different assessors scoring consistently for the same trainee performance Other terms include test-retest, inter-rater agreement Critical for validity and acceptability, as in practice trainees are scored by different assessors and over multiple occasions throughout training Tests used to measure external reliability have included: ICC: multiple types, includes agreement if stringent form used Cohen's kappa: applicable for studies using only 2 assessors ICC or Cohen $\kappa \geq 0.80$ are required for summative assessments Generalisability: type of ANOVA to estimate errors of variance (Pearson correlation is criticised for only being measures of association)

Test-retest reliability	Consistency of scores over different times for the same performance
Feasibility	Minimal time, logistic preparation, and low financial costs to administer Minimal training of assessors required
Educational catalyst	The tool is a good formative assessment for learning (generates results that allows structured feedback, deliberate practice, reflection)
Acceptability	Acceptability of the tool will be dependent on the stakeholder. For example, medical registration boards will only accept the test scores as an indicator of fitness for independent practice if high reliability and validity is demonstrated. Feasibility and educational effect are not as important. A clinical tutor accepts a formative assessment tool only if educational benefit and feasibility are high; the other properties are not as critical

ICC; Intraclass correlation coefficient

1.6.2 Statistical analysis of reliability

The most common type of reliability analysis in medical education is derived from classical test theory. However, several reviews have noted severe limitations in the statistical analyses employed by published studies.[148, 150] These have included heterogeneity of statistical tests, incorrect use of tests, and inconsistencies in methodology. This has precluded any meta-analyses of tools currently used for assessment of medical procedural skill.

In classical test theory, it is assumed that the observed score X is the sum of the true score T and any errors in measurement E, and described by the formula $X = T + E$. Reliability coefficients are derived to estimate the amount of error, which provides an indication of how accurate the observed score is to the true score. Analyses of reliability based on classical test theory have intuitive appeal, as higher reliability coefficients suggest low errors and greater consistency in observed test scores.[147]

One application of classical test theory is the use of split-half methodology for evaluation of internal reliability. This is a correlation analysis between test scores from one half of questions in a test against scores from the other half. This analysis is only valid if all questions are examining the same construct, and is most appropriate for assessment tools using a large number of MCQs. An advantage of split-half methods is identifying questions with low correlations, prompting a re-writing or removal of these questions.

Another popular classical test theory analysis for internal reliability is Cronbach's alpha (α). α is applicable for continuous data, and is in turn the general form of the Kuder-Richardson-20 test used for dichotomous data. α is calculated as a parallel test design, meaning that the coefficient is the sum of the average variance and average error of each item in the assessment tool.[151] Despite being a common test, α has been incorrectly used and its limitations under-appreciated.

Firstly, α is artificially increased when used for criteria-based assessments, and by having more items in the test.[144] α is also sensitive to individual inter-item correlations; a subset of items may weigh disproportionally on the overall average α value for the entire tool. Equally, the test is susceptible to missing data with resultant low α values. To provide a more accurate result, recommendations are to calculate standard errors of mean with 95% confidence intervals to provide information on the absolute agreement of scores.[144, 149]

In clinical practice, it is the consistency of different assessors, or external reliability, that is of greater importance than internal reliability. This is particularly true when it is a high stakes summative assessment; even in formative assessments, sufficiently high external reliability is still required so that feedback can be appropriate - assessors need to agree on ranking of performance in a procedural task, so that feedback is consistent and suitable.

Similar to internal reliability, there are several methods to analyse external reliability. In the test-retest method, consistency of assessor scoring obtained at initial testing is compared to scores obtained after retesting at a later time. This is a test of score stability, and requires a balance between retested too early (recall and learning bias) and too late (assessors may have gained experience in assessing, participated in workshops, or lost to follow up).

The choice of statistical tests is dependent on the number of assessors used for scoring. If only two assessors are used, Cohen's kappa (κ) or simple percentage comparisons have been used. Cohen's κ is a more robust test than percentage comparison, as chance agreement is taken into account in the κ statistic. For more than two assessors, the intra-class correlation coefficient (ICC) statistic or Pearson rho (ρ) product-movement correlation is applicable.

ICC is a form of analysis of variance (ANOVA) and is the preferred test. Unlike Pearson correlation, the ICC statistic takes into account not just association of test scores, but also absolute agreement between scores. Another advantage for ICC is the correction for variance based on chance alone. As a consequence, Pearson ρ sets the upper limit of ICC values in ideal situations, but calculated ICC is much lower. Studies that only use Pearson correlation therefore overstates the actual external reliability of the assessment tool.[152]

ICC itself has six statistical forms, and the choice of model will markedly change the result.[153, 154] Shrout and Fleiss demonstrated that the six ICC models produces results ranging from 0.17 to 0.91 using the same raw data.[155] It is therefore critical that the model used for analysis is stated in the methodology so that readers can appraise the validity of the ICC result. Indeed, McGraw and Wong note that Cronbach's α is simply a consistency model ICC, and is incorrectly chosen as the statistical test when analysing criteria-based tools.[154]

As an example of using ICC to analyse an UGRA assessment tool, the following factors are evaluated: is the assessed task a random selection of all UGRA procedures, and have the assessors been randomly chosen from all possible assessors (one versus two way, fixed versus random effects model); is it a single assessor, or are two or more assessors concurrently scoring the trainee (single versus average measures model); is absolute agreement between scores is required. Absolute agreement in this context denotes whether the numerical value of each item has implication; a score of 3 meaning clearly satisfactory performance has consequence over a score of 2 meaning indefinite, and has distinct consequence over a score of 1 meaning unsatisfactory performance. The most stringent model is ICC(2A,1), which is used when a single randomly chosen assessor scores a randomly chosen trainee performing a randomly selected UGRA task, with absolute agreement in scores. This absolute agreement ICC model is therefore the correct model to evaluate current workplace-based assessment tools.

Customary 'rules of thumb' for acceptable reliability thresholds have been published. For summative assessments, reliability thresholds are required to be at high levels, with recommendations that Cronbach's α should be greater than 0.90, and ICC greater than 0.75[152] or between 0.80 - 0.89 for end-of-year/rotation exams. For certification examinations, a very high stringency threshold of ICC greater than 0.90 have been advocated.[145, 147] For formative assessments, Cichetti recommended that Cronbach's α should be 0.80 - 0.89 and ICC should be 0.60 - 0.74.[152] Downing suggests similar ICC thresholds of 0.70 - 0.79. Thus, while formative assessment reliability thresholds are less than summative, they remain reasonably rigorous. This is based on the belief that it is still necessary to ensure that the trainee's performance, upon which structured feedback is provided in formative assessment, continues to be assessed consistently. Otherwise, feedback from one assessor may be unjustly negative while feedback from another assessor may be incorrectly positive despite viewing the same procedural performance.

Generalisability theory is the second major statistical technique to evaluate reliability. This theory is a form of ANOVA, and calculates the ratio of true variance (true score) to total variance (true score and errors). Generalisability is uncommon compared to analyses using classical test theory, but has been used to evaluate the pilot study of workplace-based assessments by the United Kingdom General Medical Council,[156] and colonoscopy performance.[157] In anaesthesia, generalisability has been used in a simulation-based study on paediatric anaesthesia skills,[158] and was used to validate the mini-clinical evaluation examination tool for anaesthesia procedural performance.[159]

An advantage of using generalisability over classical test theory is the ability to identify and quantify factors that contribute to decreased assessment tool reliability. In clinical studies, the two primary sources of variance are due to trainees attempting to perform the same procedure but on patients of different difficulty (case specificity), and scoring inconsistency from assessors (inter-rater error variance).[147, 160] This helps identify which areas of assessment require redesign. For example, generalisability can determine if an otherwise satisfactory trainee underperformed due to a particularly difficult patient, or if a particular assessor is consistently scoring differently from other colleagues. However, analogous to

ICC analysis, the assumptions used when performing generalisability modelling has significant effects on the resultant coefficient.[144, 161]

Another unique property of generalisability is performing reliability modelling, akin to a sample size power calculation. Based on the data set, a decision study (D-study) estimates the effect on reliability when changing the number of assessments performed versus the number of assessors required.[144] D-studies can help answer these questions: during curriculum design, what logistical resources are required to administer assessments; when designing prospective research, what is the required mix of number of assessments and number of assessors; and what is the impact on reliability if assessors receive training to improve consistency. An example D-study is shown in Figure 1.12.

One reason for reduced assessor consistency may be poorly written questions or items which lack objective clear definitions, making it difficult to score consistently. Typically, pilot testing and assessor training allows removal of ambiguous wording and clarification of item descriptions. Alternatively, a measurable endpoint is used for the task; as an example, motor fatigue while attempting an UGRA procedure is scored by a surrogate endpoint of the number of times the trainee swapped hands on the transducer or needle.

The third option to create assessment tools with inherent reliability is to use item response theory (IRT). Unlike classical test theory and generalisability models, IRT is based on study of individual items in the tool rather than scores. Each item is assessed to its probability of a correct answer versus trainee ability in an iterative process of item calibration and proficiency estimation. Discriminatory power of each item is determined by the location index, which is a graphical representation of the "item characteristic curve". More complex two- and three-parameter logistic models allows better discrimination between trainees of variable ability, but at the requirement of very large sampling sizes.[144]

Rasch modelling is a mathematical construct that employs one-parameter IRT, but has the additional requirement of model fit. Practically, during the iterative process, Rasch identifies misfit items, which may then be examined for causes of poor discrimination and amended appropriately, or may instead be excluded completely from the overall model.

A recent example of IRT and Rasch modelling in anaesthesia was the creation of a 47-item multiple choice assessment tool for sonoanatomy.[162] The MCQ was constructed using multiple iterative rounds of item content creation, piloting and revision.

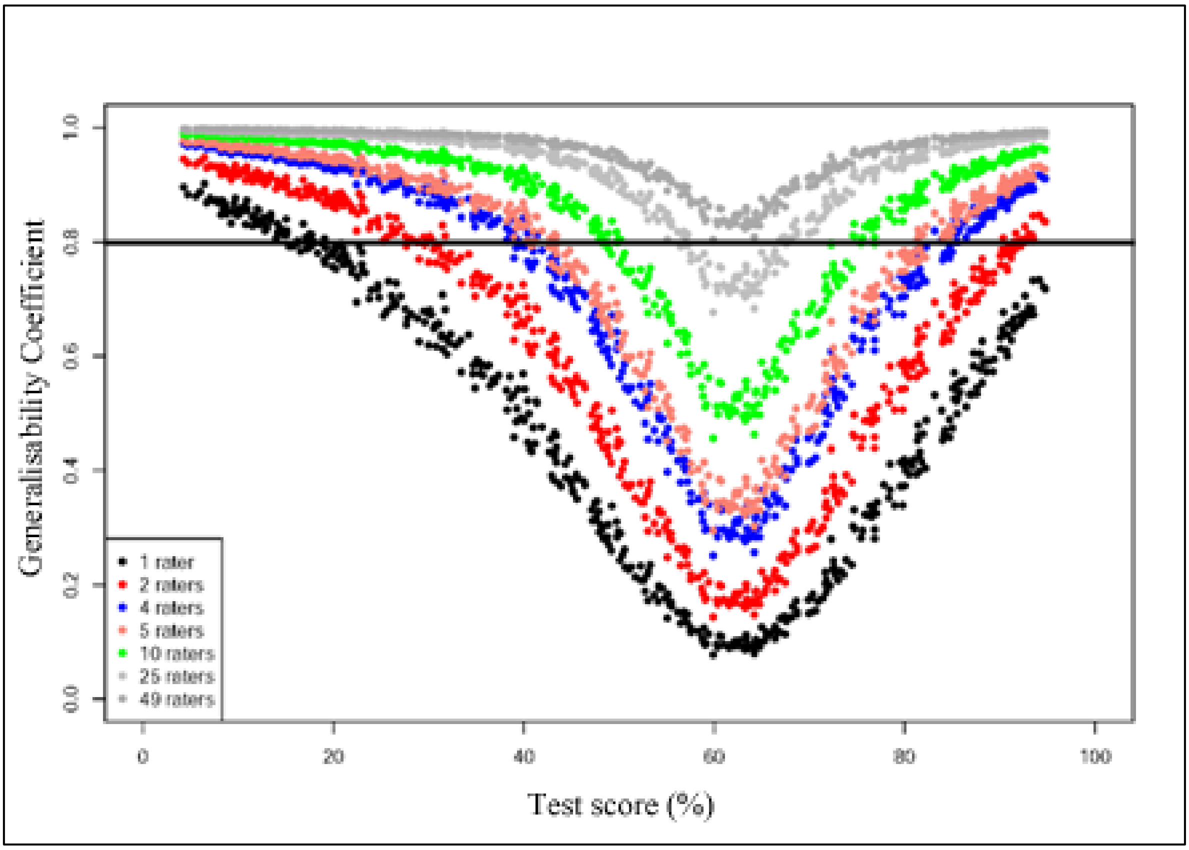

Figure 1.12. Decision study analysis. An estimated score given by a random assessor when using a specific assessment tool is plotted on the x-axis, based on the mean and standard deviation of actual scores from all assessors. The decision study is run with a variable number of assessors (1, 2, 4, 5, 10, 25, or 49 assessors) being used to simultaneously evaluate a trainee. The resultant generalisability coefficient for reliability of the assessment tool is plotted on the y-axis. This example of a Decision study used raw data from Study Two of this thesis. Acknowledgement to Dr Petra L. Graham from the Department of Statistics, Macquarie University, for assistance in construction of the graphical plot (Unpublished).

Assuming a minimum reliability coefficient of 0.80, an ideal assessment tool will be above this threshold at all test scores (excellent construct validity) and with the minimum number of assessors required (excellent inter-rater consistency). The significance would be that only a few number of assessments are required to accurately determine the true performance level of a trainee.

1.7 Assessment Tools for Regional Anaesthesia

Assessment of modern, complex medical procedural skills such as ultrasound-guided regional anaesthesia requires evaluation of disparate knowledge and skill sets. Practically, this can only be achieved using multiple types of assessment and conducted by a broad range of faculty members. This minimise biases that occur from reliance on a single assessor's opinion, basing evaluation on a single clinical performance, or assessing only a single type of UGRA procedure. On-going multiple assessment also allows tracking of progress, providing an opportunity for remediation earlier in the training program. While challenging to define and assess, non-technical skills should be included in a holistic, comprehensive approach to trainee evaluation. Assessments must be anchored on pre-defined standards of performance and behaviour.[95, 121, 143, 148, 163, 164]

In this section, studies for which the primary aim was the psychometric evaluation of an assessment tool of regional anaesthesia will be reviewed. Studies have been published for assessment of regional anaesthesia knowledge and skill sets using multiple choice questions, motion analysis, cumulative sum (cusum), visuospatial and psychomotor aptitude testing, checklists, and global rating scales (GRS). One study described the use of a checklist tool for assessing regional anaesthesia, within a larger five station objective structured clinical examination (OSCE) format used in a national anaesthesiology credentialling exam. These studies are summarised in Table 1.8.

Also included for review are tools that are intended for generic assessment of all anaesthesia procedural skills, but not previously and specifically validated for regional anaesthesia. These include global rating scales reformatted as the Direct Observation of Procedural Skills (DOPS) and the multisource feedback (MSF), the mini-Clinical evaluation examination (mini-CEX), and the Anaesthesia Non-Technical Skill (ANTS) tools.

1.7.1 Multiple Choice Questions

MCQ tests have several advantages, including the ability to assess multiple knowledge areas in the same test. Marking is rapid using computer based scanning, and large numbers of trainees are assessed at same time. MCQs can also be context-poor or context-rich. Context-poor tests are designed for assessment of factual recall without reference to clinical issues, while context-rich questions contains clinical information that trainees must consider in their decision making. Also called script-concordance questions, these context-rich MCQs help discriminate trainees on their level of clinical decision making and diagnostic reasoning.[143]

In UGRA, Woodworth et al used IRT and Rasch modelling in their aim to create a MCQ tool to assess sonoanatomy and anatomy knowledge for femoral, sciatic, and above clavicle brachial plexus blocks.[162] Using multiple rounds of content creation with different groups of expert UGRA practitioners, an initial 81-item MCQ was reduced to 50-items. These MCQs were context-poor questions requiring factual recall of knowledge. MCQs were computer based, and included photographs, diagrams, and/or ultrasound imagery.

Validation of the remaining MCQs was performed with IRT analysis, using the test results obtained from 90 participants of different experience and seniority levels who answered at least 80% of all MCQs. 3 items were identified with Rasch modelling as having poor fit and were removed, resulting in a 47-item MCQ assessment tool with face, content, and construct validity.

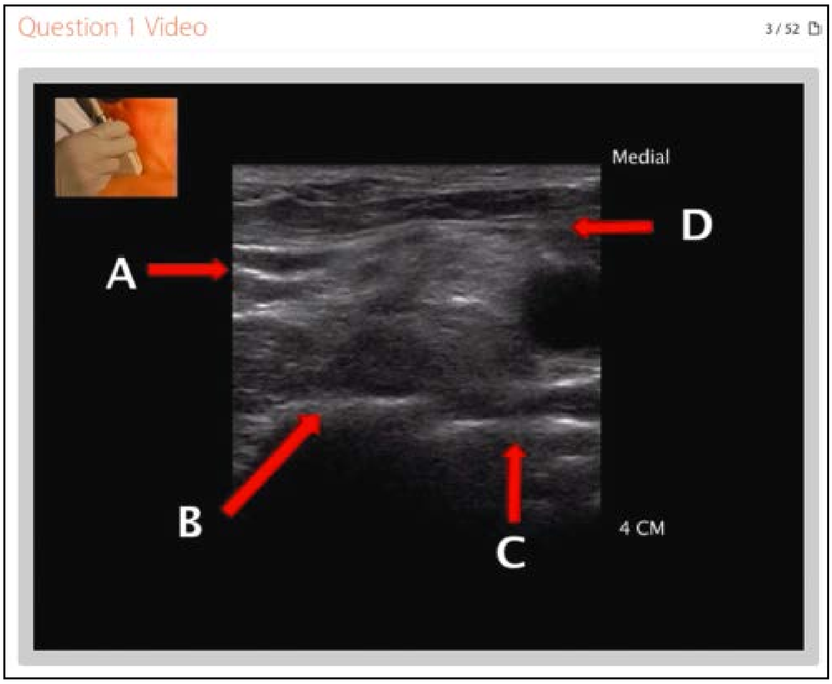

Figure 1.13. Example MCQ. Reproduced from Woodworth et al.[162]

1.7.2 Motion analysis

Motion analysis uses a radiofrequency tracking system to monitor the dexterity of proceduralists. This is achieved through tracking of sensor-embedded gloves worn by the proceduralist. The majority of motion analysis studies used the Imperial College Surgical Assessment Device (ICSAD), which has been previously used to measure surgical skill,[165] and more recently central venous cannulation performance.[166, 167] Concurrent validity with the OSATS test, considered to be the gold standard assessment of surgical procedural performance, was only moderate in the surgical studies.[165] Three objective endpoints are measured: time taken to perform a task, number of movements, and economy of motion ("path length", equating to the distance travelled by the hands). These can be numerical or graphically plotted, as illustrated in Figure 1.14.

In the one study in UGRA by Chin et al, ICSAD was evaluated for construct validity between novice and expert UGRA proceduralists when performing an ultrasound-guided supraclavicular brachial plexus block in a clinical setting.[168] This showed that consultants were significantly superior to trainee anaesthetists in all three objective measures of dexterity. Fellows undertaking a 12-month subspecialisation in UGRA were found to have performance nearing that to consultants by the end of their training, improving significantly from the beginning of their fellowship year.

As a secondary aim, ICSAD performance was concurrently validated with a combined checklist and GRS assessment tool. This combined tool was modified from a previous checklist and GRS used in another ultrasound-guided supraclavicular plexus block study,[169] namely with addition of 10 extra items in the checklist and deletion of one item from the GRS. Inter-rater reliability was measured by Pearson ρ correlation, and concurrent validation was analysed by Spearman rank correlation. While the authors reported correlation coefficients of up to 0.97, true reliability will be below this upper bound.

1.7.3 Cumulative sum

Cusum is a sequential analysis statistical method initially used as a tool for quality assurance during manufacturing of goods on a factory floor. If process control was noted to be trending away from a pre-determined measure of quality, this early recognition allowed an opportunity for remediation. Using the same principle, a trainee's learning curve can be plotted over successive attempts, and is used to track performance changes towards either achieving or moving away from an agreed competency criterion.[170]

In the three studies performed on regional anaesthesia performance using cusum analysis, the sequential probability ratio test form of cusum was used. This requires an *a priori* decision of the acceptable α error (risk of mislabelling a trainee as incompetent) and β error (risk of mislabelling an incompetent trainee as competent), as well as acceptable (p_0) and unacceptable (p_1) failure rates. This generates statistical thresholds for either achieving or not attaining competency over time, respectively denoted as the h_0 lower and h_1 upper decision thresholds. Determination of success or failure for each attempt is a binary outcome and uses an objective endpoint. Performance that is trending towards competency is graphically represented as a decreasing cusum plot that finally crosses the h_0 lower threshold, while worsening performance causes the plot to increase towards the h_1 upper threshold (Figure 1.15).[121]

Multiple disadvantages for using cusum in medical education have been raised. These include: requirement for statistical expertise; logistical complexity as multiple attempts are required to construct the cusum plot; case specificity is not considered, thus differences in patient difficulty are unaccounted; and is useful only for part task technical performance but not of the entire UGRA procedure. More seriously, the cusum formula is a statistical method that disproportionately weighs previous performances, which may not be reflective of current levels of competency.[170]

Kestin developed cusum plots for four different anaesthesia procedures, including arterial and central venous cannulation, obstetric epidural anaesthesia, and spinal anaesthesia.[171] Success for epidurals was defined as establishment of adequate analgesia or anaesthesia, and success for intrathecal anaesthesia was ability to aspirate cerebrospinal fluid through the spinal needle. De Oliveira Filho performed a similar study, defining success at epidural insertion as loss of resistance to saline, and spinal insertion as flow of cerebrospinal fluid.[172]

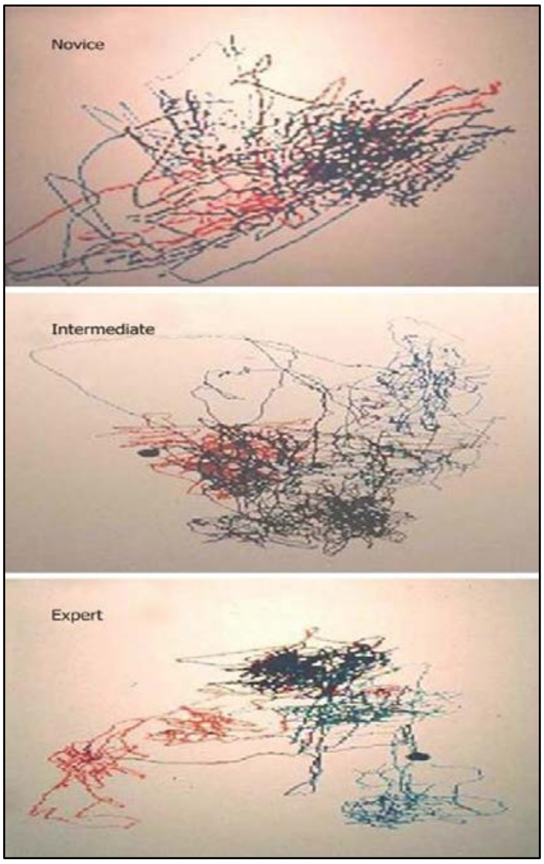

Figure 1.14. Hand motion analysis. Graphical representation of novice versus intermediate versus expert proceduralist path lengths. Acknowledgment to Dr Sanjib Adhikary, MD, Pennsylvania State University, for providing this example.

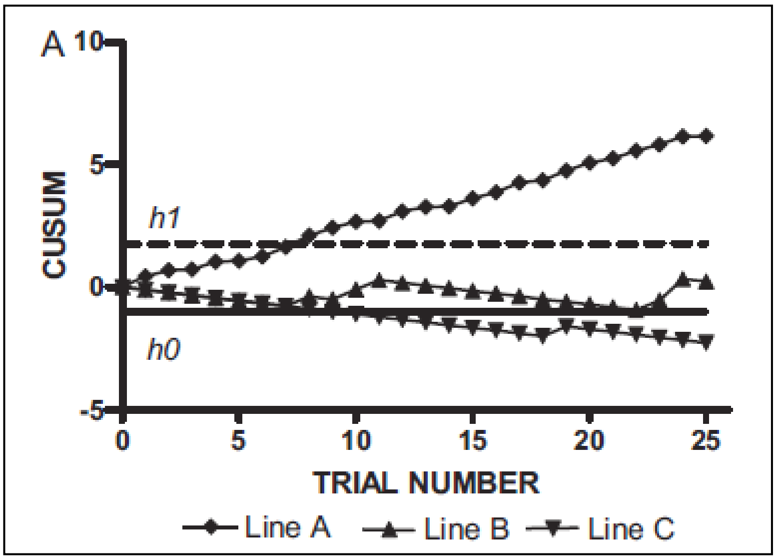

Figure 1.15. Cumulative sum analysis. Average cusum lines for novices using real-time ultrasound for needle visualisation and guidance to a target inside a meat phantom. Line A represents novices who did not reach proficiency and remained above h_1 decision threshold, Line B represents novices for which proficiency could not be statistically determined, and Line C represents novices who did achieve proficiency and crossed below the h_0 decision threshold. Reproduced from De Oliveira Filho et al.[113]

In UGRA, De Oliveira Filho et al performed a study using novices to manipulate a needle under real-time ultrasound-guidance to a pork tendon target embedded in a bovine *in vitro* phantom. The results of this study showed that 109 attempts were required for 95% of novices to reach competency, defined as a score < 2 on an image quality scale rated by 2 blinded assessors.[113]

Barrington et al repeated this study on UGRA needle skills, using the sciatic nerve of a fresh frozen human cadaver as the target, and defining competency as a score < 3 on a combined image quality and transducer stability scale rated by a blinded assessor.[114] Crucially, the novices in De Olivera Filho et al's study underwent discovery learning, while the novices in Barrington et al's study received deliberate practice and feedback from an expert supervisor after each needling attempt. In this study, novices attained statistical competency after 28 attempts. In all four studies, individual trainees became competent in regional anaesthesia procedures at a highly variable rate (Figure 1.16).

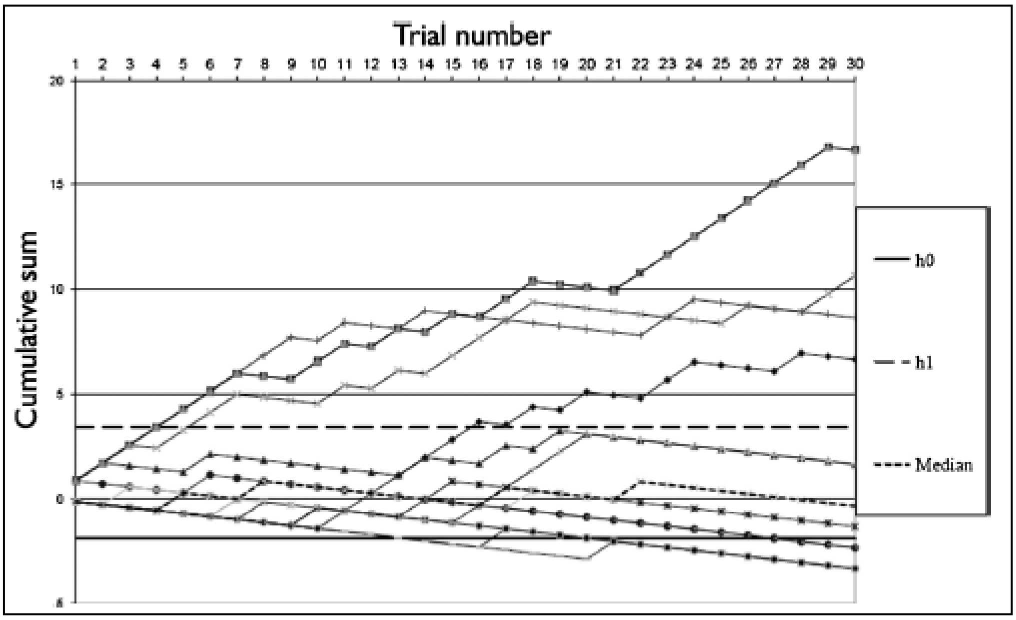

Figure 1.16. Individual cusum plots. Cusum plots of 15 novices attempting a needle UGRA task, illustrating wide individual variability in their speed of learning and number of attempts before reaching competency. 5 trainees did not reach competency by 30 attempts; a further 4 were indeterminate, while only 6 trainees achieved competency in the task. Reproduced from Barrington et al.[114]

1.7.4 Visuospatial and psychomotor ability testing

The Cattell-Horn-Carroll theory is a taxonomic system that divides human general intelligence *g* into multiple "broad stratum" cognitive abilities such as visuospatial and psychomotor ability. These abilities are composed of multiple "narrow stratum" factors, with each factor explaining a specific attribute of the higher stratum (Figure 1.17). Current neuro-psychometric tests measure these narrow stratum factors, necessitating a battery of tests to holistically describe an individual's level of visuospatial or psychomotor ability.[173, 174]

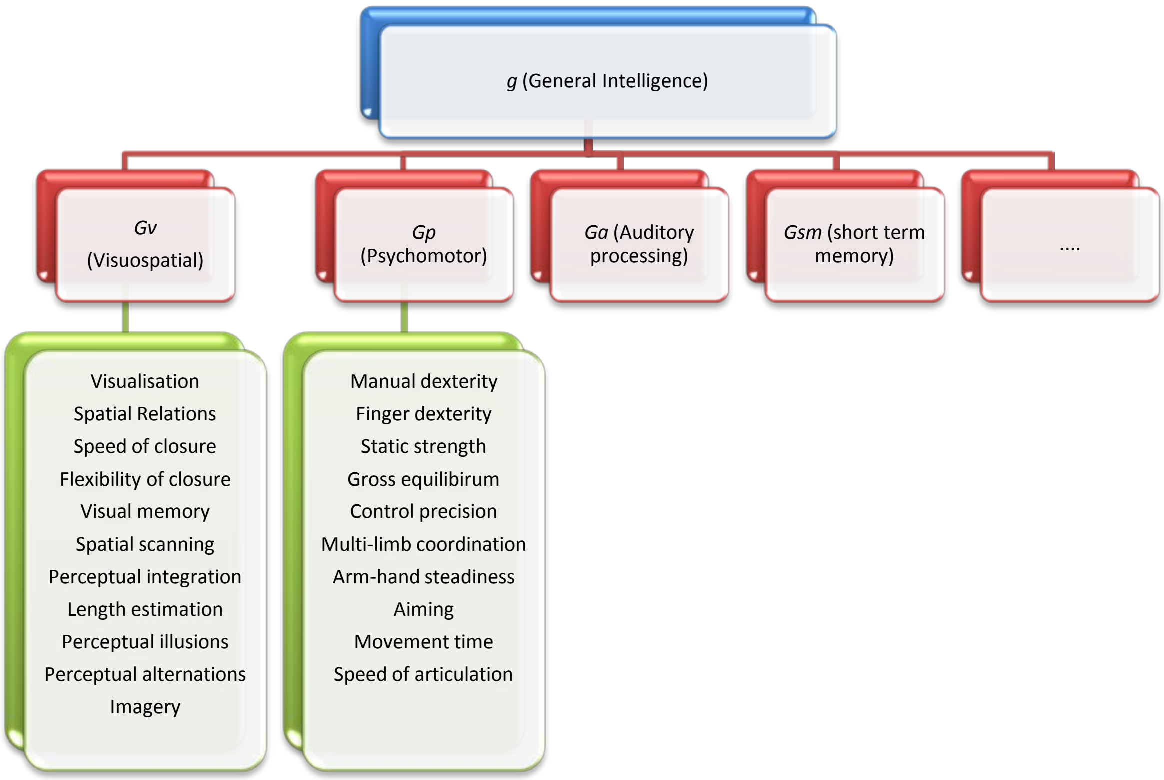

Figure 1.17. Cattell-Horn-Carroll model of human cognitive ability, version 2. Visuospatial and psychomotor broad abilities (red colour) and their component narrow stratum factors (green colour). Only 4 out of 16 broad stratum abilities are illustrated in this figure. Other broad abilities include fluid reasoning, comprehension-knowledge, long term memory, cognitive processing speed, decision and reaction speed, reading and writing, quantitative knowledge, general domain level knowledge, tactile, kinesthetic, olfactory, and psychomotor speed; these broad abilities and their narrow stratum factors were omitted for clarity. Adapted from McGrew,[173] Schneider and McGrew.[174]

Visuospatial ability itself is defined as the human cognitive capacity to generate, retain, retrieve, manipulate and process visual information.[173] Psychomotor ability influences the capacity to precisely control objects via bimanual dexterity, and incorporates hand-eye coordination, reaction speed and accuracy.[175] In complex spatial tasks, innate differences in visuospatial ability was found to contribute 40% of inter-individual learning speed and performance.[176]

The premise that some individuals are "gifted" to perform well in procedural skills is the basis for using neuro-psychometric tests to screen potential trainees. This has been incorporated into the enrolment process for dentistry training positions in some universities.[177] This stance is supported by studies in laparoscopic surgery, general surgery, interventional radiology, anatomy, and endoscopy procedural tasks, where there is considerable evidence for correlation of visuospatial and psychomotor ability with proficiency and efficiency.[178-190] When attempting to transfer from simple to complex procedural tasks, novices with higher visuospatial ability level do so with greater proficiency than their lower ability peers.[183, 184] In anaesthesia, Dashfield and Smith found that psychomotor ability correlated with initial fibreoptic bronchoscopy performance in airway novices.[191]

However, evidence also exists to suggest that psychometric testing alone is insufficient for eventual competency. For example, in Dashfield et al's study, later fibreoptic bronchoscopy performance did not correlate with psychomotor test results. Other studies have similarly found that further training negates some of the advantages of innate ability,[175, 192] while studies recruiting experts found that their psychometric test scores do not correlate with competent performance.[193, 194] These suggest that competency in medical procedural skills is a combination of innate ability as well as cognitive constructs, which can be taught or learnt through experience. Thus, the value of psychometric testing may not be to exclude low ability trainees, but in recognising that they will require different educational strategies to assist them to attain competency. This premise has not been explored in UGRA research.

There are two studies of psychometric testing for UGRA performance. Smith et al recruited anaesthesia trainees and performed visuospatial (Block Design, subset of the Weschler Adult Intelligence Scale; Digit Symbol Substitution; Trails Test; Pelli-Robson Contrast Acuity) and psychomotor ability tests (Zig-Zag; Purdue Pegboard; Crawford Small Part Dexterity; Semmes-Weinstein Sensory).[195] All trainees performed a needle UGRA task on a bench top *in vitro* model, using time and image quality endpoints. The significant finding was that only the Block Design visuospatial test correlated with UGRA task performance. Psychomotor ability was not associated with performance.

Shafqat et al used medical students in their study investigating visuospatial ability (Mental Rotations Test; Group Embedded Figures; Alice Heim Group Ability) and the effect of anxiety state and cognitive ability on performance of a needle UGRA task using a turkey breast model.[196] The results showed that only Mental Rotations correlated with higher technical performance and lower error rates, while higher anxiety levels were an adverse factor.

There are two difficulties in interpreting existing studies on visuospatial and psychomotor ability and medical procedure performance. Current psychometric tests are designed to assess individual narrow stratum factors. However, each medical procedural skill will activate a subset of all narrow factors, and different combinations of factors is likely for different tasks even in the one procedure.[197] Study design is therefore critical in choosing the appropriate psychometric test for the procedural task being evaluated. Secondly, testing is not standardised, and multiple tests are available for the same factor. For the visualisation factor in visuospatial ability, the Vandeberg and Kuse Mental Rotations Test-A is highly regarded with extensive population data (Figure 1.18).[198, 199] However, the lack of consistency in testing of other factors complicates any direct comparisons between studies.

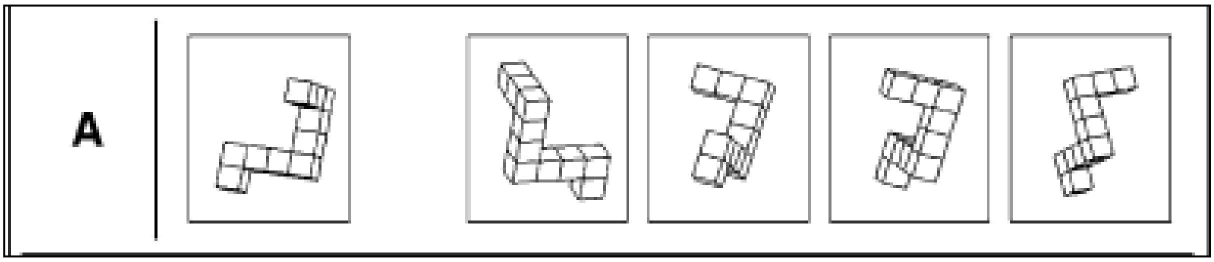

Figure 1.18. Mental Rotations Test-A. Example question of assessing spatial visualisation as part of visuospatial ability testing. This tests the ability to perceive three dimensional objects from two dimensional representations. Starting with the image on the left, the candidate has to choose the two correctly rotated options from the four images on the right. Reproduced from Peters et al.[198]

1.7.5 Checklists and Global Rating Scales

Checklists are designed to assess component skills of an entire procedure. Identification of the core skills are determined informally through consensus opinions, formally through a Delphi process, or systematically using cognitive task analysis. Each checklist is designed for a specific procedure. Scoring is usually dichotomous (satisfactory/unsatisfactory, or pass/fail).[121] During pilot and validation of new checklists, psychometric analysis of individual items allows identification of questions with poor reliability. One example in anaesthesia was the creation and refinement of a tracheal intubation performance checklist.[200]

Checklists are widely used to assess procedural skills in basic surgical skills, general laparoscopic tasks, endoscopy, and for specialty-specific tasks in vascular, cardiothoracic, and colorectal surgery.[148] In anaesthesia, Fehr et al used 8- and 10-item checklists to assess acute paediatric anaesthesia performance in a simulation centre over 10 scenarios, including management of bronchospasm, malignant hyperthermia, laryngospasm, the septic patient, airway foreign body, neonate resuscitation and postoperative apnoea.[158] Checklists were also used to assess anaesthesia skills such as preoperative assessment, preparation and induction by medical students,[201] and for central venous cannulation.[202] Murray et al used checklists to assess performance in acute intraoperative emergencies in a simulation setting, over several studies.[203-205]

Checklists provide a rich source for formative feedback as weaker component skills within a whole performance can be identified. Contrariwise, checklists reward thoroughness in trainees but not necessarily discriminate between experts, as the latter group tend to employ short-cuts in technique.[121]

Unlike checklists, global rating scales (GRS) is a generic tool that can be used to assess different procedures. GRS also differs from checklists in including assessment of both technical and non-technical aspects of performance. To do so, GRS item descriptions are non-specific behavioural anchors, and the scoring scales are Likert-based. This is useful for domains where dichotomous checklists are unsuitable, such as communication skills, team work, insight, list management and resource utilisation/efficiency.

The main disadvantage of the qualitative nature of the GRS includes the halo effect, where the general impression of the trainee may influence the scoring of individual items. Similarly, scale limitation may occur with assessors marking in a small range of Likert scores instead of using the entire scale.[121] This has led to a belief that dichotomous checklists are objective, while the subjectivity inherent in GRS is seen as a shortcoming. However, in multiple clinical studies, and repeated in a recent meta-analysis, both GRS and checklists appear to be at least equally reliable.[201, 206-208]

The original form of the GRS was a 7-item, 5-point Likert scale tool. This was introduced concurrently with task-specific checklists designed for the OSATS surgical skills assessment by Martin et al in 1997.[209] Since then, the GRS has been used unchanged, modified for specific studies, or used as the basis for GRS-like tools.

Combining a checklist with a GRS creates a more comprehensive assessment tool. Each component gathers different data on performance, and makes available the advantages of both types of tools. Indeed, Bould et al in their editorial suggests that the gold standard for clinical procedural assessment is a tool with a combined checklist and GRS.[121] There is thus only one study that described the creation of a stand-alone checklist tool for regional anaesthesia assessment, while all other studies have a combined checklist and GRS.

Sivarajan et al created a 61-item lumbar epidural checklist using consensus opinions from experts.[210] They defined 8 "critical items" such as accidental dural puncture which was deemed an automatic unsatisfactory performance for the entire checklist. Identifying critical items creates a weighted checklist that prevents trainees receiving a high total score, despite committing clinically significant errors or not completing essential tasks.

For initial piloting, assessors rated 8 videotaped lumbar epidural insertions. Each assessor was provided an instruction book with item descriptions to improve consistency in scoring. Moderate-high reliability was confirmed with Cohen's κ, but neither construct validity or feasibility was evaluated.

Friedman et al[211] designed and evaluated a 27-item checklist for obstetric lumbar epidural analgesia, combined with a 7-item GRS modified from the original OSATS. A cohort of six anaesthesia trainees were videotaped at the beginning, middle, and end of their obstetric anaesthesia rotation. These experience levels corresponded to novice (< 30 attempts), intermediate (31 - 90 attempts), and experienced (> 90 attempts) trainees. 21 videos of epidural insertions by these trainees were collected over a 6 month period.

Psychometric analysis for external reliability was Cohen's κ, construct validity using repeated measures ANOVA, and consistency of scoring between the checklist and GRS with Cohen's κ. Only 3 individual items of the checklist, and 1 item of the GRS, scored at a high stakes summative threshold of ≥ 0.75. This suggests poorly interpreted items or that assessor training was insufficient to obtain consistent scoring. In the paper, training of assessors was not elaborated, but the authors did note that some assessors exhibited scale limitation bias. However, total scores of both checklist and GRS had κ values of 0.90 and 0.87, respectively. Total scores also significantly improved with greater experience. This suggests that assessors can still form a reliable, holistic opinion of a trainee's overall performance despite being inconsistent with individual item scoring.

Naik et al used a Delphi technique with 10 expert anaesthetists at the authors' institution, and over 4 iterative rounds of review, reached a consensus on a 20-item checklist for a surface landmark-guided interscalene brachial plexus block.[169] This was paired with a GRS used previously for fibreoptic intubation assessment.[212] 10 novice trainees (< 10 previous blocks) and 10 more experienced trainees (> 10 previous blocks) were recruited.

The results did not clarify how many videos were collected for analysis. External reliability was 0.74 and 0.85 for the checklist and GRS respectively, but true reliability will be lower as these results were evaluated using Pearson ρ correlation. Construct validity was confirmed by t tests of checklist and GRS total scores, with significant differences between novice and more experienced trainee scores. Feasibility was not evaluated.

Sultan et al created a checklist for ultrasound-guided axillary plexus blocks derived initially from a 35-item checklist constructed using cognitive task analysis. This was expanded to the final 63-item checklist through expert consensus,[213] and combined with a GRS to create the assessment tool. 15 anaesthetists of different experience levels were recruited to represent novice, intermediate, and expert UGRA proceduralists. Each anaesthetist performed 2 axillary plexus blocks, and were scored by 2 blinded assessors. For this study, external reliability was evaluated by Cronbach's α, a consistency model ICC(C,*k*). Construct validity was confirmed for both checklist and GRS, but feasibility was not evaluated.

Cheung et al adopted a Delphi technique to construct a checklist and GRS suitable for all UGRA blocks.[214] This process recruited 18 experts in 6 different institutions, and had 3 rounds of iterative discussion and consensus to establish a 22-item checklist and 9-item GRS. Two subsequent studies have since been published to evaluate the tool's psychometric properties, both in 2014: Study Three of this thesis,[215] and a study by Burckett-St Laurent et al.[216]

In this latter study, the authors aimed to demonstrate construct validity in a simulation-centre based setting, concurrent validity using the tool in a clinical setting, and external reliability. 20 anaesthetists, defined as novice (< 50 previous UGRA blocks) and experienced (> 50 blocks) performed a supraclavicular UGRA block each in a patient and in a bench top model. Construct validity was tested with t tests, concurrent validity by Pearson ρ correlation, and reliability by two-ways average measures ICC. While not specifically explained, total scores of both checklist and GRS were presumably used for analysis. Individual item data was not reported. Feasibility of the assessment tool was also not reported.

Construct validity was demonstrated for the GRS, but the checklist failed to discriminate between novices and experienced anaesthetists on the simulator model. There was significant correlation of participants' checklist and GRS scores when performing the block in the clinical and in the simulation setting. Reliability of the checklist and GRS was > 0.75 in the clinical setting, but failed to reach this high stakes threshold for the bench top model.

Interestingly, the authors contend that their simulator is a high fidelity model, and that failure of the checklist to demonstrate reliability and construct validity is most likely due to deficiencies in the checklist. However, this model is actually a low fidelity gelatin-based model (Blue Phantom, Redmond, Washington). The quality of the ultrasound images of the musculoskeletal and neural structures is unrealistic compared to *in vivo* images (Figure 1.19). Use of a proper high fidelity model may change the results of this study, as observed in other studies comparing transfer of skills from bench top to clinical, using a combined checklist and GRS.[217]

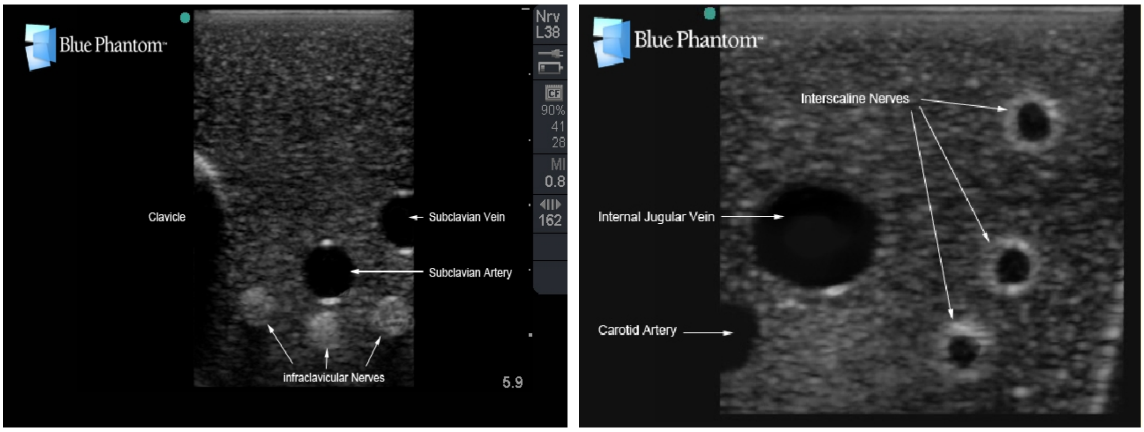

Figure 1.19. Blue Phantom UGRA model. Left is representing the infraclavicular brachial plexus, and the right is for the interscalene brachial plexus. No image for the supraclavicular brachial plexus is publicly available. Images were downloaded from the Blue Phantom website (http://www.bluephantom.com/product/Regional-Anesthesia-and-Ultrasound-Central-Line-Training-Model_NEW.aspx?cid=428 accessed June 17, 2016)

1.7.6 Objective Structured Clinical Examination and Objective Structured Assessment of Technical Skills

The OSCE was originally described by Harden et al in their 1975 and 1979 papers, and introduced the format of candidates rotating through multiple stations.[218, 219] A specific knowledge base, clinical skill set, or procedural task is tested at each station under time constraints. The content at each station is highly modifiable: models may be patients or *in vitro* bench top tasks, the patients may be real or played by actors, setting could be in a simulation or clinical environment; stations may cover history taking, examination, interpretation of investigation results, performing a procedure, or testing knowledge; breadth of examination could be the entire national anaesthesia curriculum[220, 221] or designed to test specific skills, such as transfusion practices in anaesthesia.[222]

The flexibility of the OSCE format has popularised this method of administering examinations, and has been investigated in more than 1600 published studies.[223] The disadvantages are organisational, logistics, and cost. A large pool of trained assessors is necessary, along with equipment, organising availability of patients, venue hire and flow of trainees through stations.

The validity and reliability of any particular OSCE is itself dependent on the assessment tools used at individual stations. These tools are criterion-based for each station, and can be MCQs, written short answers, and task-specific checklists with or without a GRS. The OSATS, introduced by Martin et al in 1997,[209] and psychometrically evaluated by Goff et al,[224] is a type of OSCE designed for surgical skills assessment. Overall reliability and validity of the OSCE/OSATS format is improved by having many stations and multiple assessors. This reduces any limitations caused by case- or content-specificity at any particular station, and minimises inter-rater scoring inconsistency due to assessor bias or lack of assessor training.[225, 226]

The Israeli Anaesthesiology National Board uses an OSCE for trainee examinations. Ben-Menachem et al described the development and psychometric evaluation of the mandatory regional anaesthesia station in the 5-station OSCE.[227] Content and face validity was established through a Delphi process using 10 regional anaesthesia experts, resulting in 8 possible scenarios (interscalene, axillary, sciatic, popliteal, ankle, cervical, epidural blocks, and regional anaesthesia for awake fibreoptic bronchoscopy; all were non-UGRA techniques). Each scenario was scored by 2 assessors, using a 12- to 20-item dichotomous checklist and 1-item GRS. The checklist was weighted and contained 2 or 3 critical items. Trainees passed the station if total checklist score was > 70%, completed all critical items, and scored > 50% on the GRS.

The psychometric analysis was explained in the authors' earlier papers.[228, 229] Pearson ρ correlation was performed on the percentage rate of both assessors scoring in unison for all items in the checklist, only the critical items in the checklist, and for the GRS. This was reported as 0.89, 0.86, and 076 respectively, and explained as the inter-rater reliability. Ideally, Cohen's κ would have been a more robust test and would account for chance agreement. There was no significant correlation between OSCE scores and performance at the *viva voce* component of the National Board examinations.

Table 1.8. Summary of clinical studies investigating assessment tools for regional anaesthesia procedural performance. Only includes studies where the primary aim was design and/or validation of tools assessing regional anaesthesia knowledge, skill sets, or behaviours.

Reference	Description	Comments
Multiple Choice Questions		
Woodworth (2015)[162]	47-items, sonoanatomy relevant to femoral/ sciatic/above clavicle brachial plexus blocks	Constructed with item response theory and Rasch modelling
Motion analysis		
Chin (2011)[168]	Imperial College Surgical Assessment Device. Construct validation. Secondary aim of concurrent validation with a modified 30-item checklist and 7-item GRS (original 20-item checklist and 8-item GRS from Naik[169])	Primary aim of construct validation achieved with discrimination between novices and experts, and between trainees at beginning to end of 12-months' fellowship training
Cumulative Sum analysis		
Kestin (1995)[171]	Investigated epidural insertion, spinal anaesthesia, central venous cannulation, and arterial cannulation	Learning curves are highly individual and variable
De Oliveira Filho (2002)[172]	Investigated epidural insertion, spinal anaesthesia, peripheral vascular cannulation, and tracheal intubation	64% can perform spinals competently, but required exposure range is 16–56 attempts.
De Oliveira Filho (2008)[113]	UGRA needle guidance skills, bovine phantom with pork tendon target. Competency based on image quality score	109 attempts required by novices for UGRA needle task; discovery learning process
Barrington (2012)[114]	UGRA needle guidance skills, fresh cadaver model. Competency based on combined image quality and transducer stability score	28 attempts required for UGRA needle task when deliberate practice teaching is used

Reference	Description	Comments
Visuospatial/Psychomotor factors		
Smith (2012)[195]	UGRA needle guidance task. 4 visuospatial and 4 psychomotor tests used on anaesthesia trainees	Only the Block Design, which tests spatial relations and speed of closure, was significant
Shafqat (2015)[196]	UGRA needle guidance task. 3 visuospatial tests used on medical students. Anxiety state and general cognitive abilities also tested	Only the Mental Rotations test, which tests spatial visualisation, was significant. High anxiety levels decrease performance
Checklist only		
Sivarajan (1981)[210]	61-item checklist specific for lumbar epidurals. Weighted with pass/fail items	Moderate reliability. Construct validity, feasibility not evaluated
Combined Checklist and Global Rating Scale		
Friedman (2006)[211]	27-item checklist specific for epidurals Combined with 7-item modified GRS used previously for surgical skills assessment (Martin et al[209])	Confirmed construct validity and reliability of overall score, but individual item reliability was poor. Feasibility not tested
Naik (2007)[169]	20-item checklist specific for UGRA supraclavicular block. Combined with 8-item GRS used previously for fibreoptic intubation (Naik[212])	Delphi technique for checklist creation. Construct validity confirmed, but reliability measured with Pearson ρ
Sultan (2012)[213]	63-item checklist specific for UGRA axillary block. Combined with 8-item GRS used previously for fibreoptic intubation (Naik[212])	Checklist designed in part using cognitive task analysis. Construct validity confirmed
Cheung (2012)[214]	Creation of 22-item checklist, and modified 9-item GRS (Martin et al[209]), specific for UGRA	Delphi technique used 18 experts in 6 different institutions for content validity. Study did not evaluate psychometrics
Burckett-St Laurent (2014)[216]	Psychometric evaluation of the UGRA checklist and GRS	Construct validation, concurrent validation between bench top and clinical setting. Reliability measured with ICC

Reference	Description	Comments
Objective Structured Clinical Examination		
Ben-Menachem (2011)[227]	Regional anaesthesia is mandatory station in the 5-station OSCE. 8 scripted scenarios. Scenarios assessed by 12- to 20-item checklist and 1-item GRS	Delphi technique for checklist creation. Reliability measured with Pearson ρ. Convergent validity of OSCE scores with oral examination scores not significantly correlated

1.7.7 Derived tools from global rating scales

A necessary component of competency-based postgraduate medical education is concurrent use of criteria-anchored workplace-based assessments (WBA). This has resulted in the introduction of several WBA tools constructed with Likert-based scales similar to the GRS. To allow WBAs to be used in formative assessment, free text sections are also common to allow feedback and discussion between trainee and supervisor. Two WBAs with GRS scales are the Direct Observation of Procedural Skills (DOPS) and the Multisource Feedback (MSF, also known as 360° assessment) tools.

The MSF is a combined narrative and Likert-based feedback given by non-anaesthetic staff, such as nurses, surgical colleagues and theatre assistants, and sometimes also by patients. MSF assesses professional behaviours and non-technical skills including teamwork, leadership, and interpersonal communication.[164] Given the subjectivity of responses, it is estimated that consistency of feedback is achieved only when more than 10 staff or more than 50 patient MSF forms are returned.[163] As a holistic assessment of non-technical skills, there are no publications using MSF for regional anaesthesia.

DOPS was first piloted in 2002 by the United Kingdom General Medical Council and Royal Colleges of Physicians to assess procedural performance in postgraduate specialist trainees.[164] DOPS was subsequently introduced into the UK Foundation Programme, which is a 2 year hospital-based training curriculum for all junior doctors to bridge the time between medical school graduation and entry into postgraduate specialist training. Confusingly, DOPS may be a generic tool (used for all procedures) or be task-specific; some were intended for formative assessment but others are used for summative pass/fail or accreditation barriers. The existing literature does not distinguish between these competing goals, even in the most recent systematic review.[230]

The largest data set for analysing the DOPS tool is based on the UK experiences. Wilkinson et al performed a study on the initial pilot.[156] They collected 118 completed DOPS forms from 59 trainees, who were assessed with both a generic DOPS form (Figure 1.20) and a task-specific DOPS form (Figure 1.21) on one of the four procedures that were evaluated: cardiac catheterisation, renal biopsy, neurophysiological study, and endoscopy. Anaesthesia trainees and anaesthesia procedures were not represented. Analysis of reliability used generalisability theory.

DIRECT OBSERVATION OF PROCEDURAL SKILLS (DOPS) :
Generic Form for assessment of SpR's core procedural skills

Assessor's Code: A - ☐☐☐☐ Date (DD/MM/YY): ☐☐ / ☐☐ / ☐☐ SpR Code: T - ☐☐☐☐

Year of training: ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

Procedure being observed: (CAPITALS)

Indication for procedure/diagnosis: (CAPITALS)

Please mark one of the circles for each component of the exercise on a scale of 1 (extremely poor) to 9 (extremely good). For the purpose of the study a score of 1-3 would be considered unsatisfactory, 4-6 satisfactory and 7-9 would be considered above that expected, for a trainee at the same stage of training and level of experience. Please note that your scoring should reflect the performance of the SpR against that which you would reasonably expect at their stage of training and level of experience

Component		POOR			NEITHER POOR NOR GOOD			GOOD	
1.Indications for procedure	○ Not observed or applicable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
2. Obtaining informed consent	○ Not observed or applicable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
3. Appropriate analgesia or Safe Sedation	○ Not observed or applicable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
4.Technical ability	○ Not observed or applicable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
5. Professionalism and consideration for patient including during the procedure	○ Not observed or applicable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
6. Clinical judgement	○ Not observed or applicable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
7.Awareness and management of complications	○ Not observed or applicable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
8.Interpreting diagnostic information	○ Not observed or applicable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
9.Drawing up an appropriate management plan	○ Not observed or applicable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
10.Counselling and communication of results to patient/relatives	○ Not observed or applicable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
11. OVERALL CLINICAL COMPETENCE PERFORMING PROCEDURE	○ Not observed or applicable	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9

12. Comments on trainee's performance on this occasion (BLOCK CAPITALS PLEASE)

Figure 1.20. Generic DOPS form. As used for initial 2002 UK General Medical Council pilot. Reproduced from Wilkinson et al.[156]

DIRECT OBSERVATION OF PROCEDURAL SKILLS (DOPS) :

Additional section for Cardiac catheterisation

Assessor's Code A - [ ][ ][ ][ ]   **Date (DD/MM/YY)** [ ][ ] / [ ][ ] / [ ][ ]   **SpR Code** T - [ ][ ][ ][ ]

Please mark one of the circles for each component of the exercise on a scale of 1 (extremely poor) to 9 (extremely good). For the purpose of the study a score of 1-3 would be considered unsatisfactory, 4-6 satisfactory and 7-9 would be considered above that expected, for a trainee at the same stage of training and level of experience. Please note that your scoring should reflect the performance of the SpR against that which you would reasonably expect at their stage of training and level of experience

1. Aseptic technique

○ Not observed or applicable ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

POOR — NEITHER POOR NOR GOOD — GOOD

2. Gaining vascular access

○ Not observed or applicable ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

3. Selection of appropriate catheters and their safe and effective manipulation

○ Not observed or applicable ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

4. Acquisition of images

○ Not observed or applicable ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

5. Consideration of radiation exposure to patient, staff and self.

○ Not observed or applicable ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

6. Sheath removal and groin management

○ Not observed or applicable ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9

7. As a consultant assessor how long did the following take with DOPs:

Adapted DOPS time: Observing [ ][ ] mins Providing feedback [ ][ ] mins

8. Assessor satisfaction with DOPs documentation and process as a method of assessing SpRs

Low ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 **High**

9. SpR satisfaction with DOPS documentation and process as a method for assessing SpRs

Low ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 **High**

10 Comments by assessor on DOPS documentation and process (BLOCK CAPITALS PLEASE)

11 Comments by SpR on DOPS documentation and process (BLOCK CAPITALS PLEASE)

Figure 1.21. Cardiac catheterisation task-specific DOPS form. An example of the task-specific DOPS forms used in the initial 2002 pilot. Reproduced from Wilkinson et al.[156]

Based on a Decision study, the authors recommended that at least 6 DOPS, scored by 3 separate assessors who supervised 2 DOPS each, were required for adequate reliability. This in turn informed the UK Foundation Programme, which required their trainees to undertake a minimum of 6 DOPS over 2 years.

Mitchell et al undertook a retrospective analysis of the 75,580 WBAs collected from 1646 trainees over a 4 year period in the North Western deanery.[231] Records were used to identify doctors who were reported as trainees in difficulty, defined as when problems in professional competency, irrespective of cause, affected their ability to work as junior doctors and this concern required unusual escalation to the deanery director to take or recommend action. The primary aim of this study was construct validity: to determine if WBA scores were associated with trainee underperformance, and if WBA scores can identify and predict the trainee with difficulties. Of the four WBAs evaluated, the mini-CEX and the Case Based Discussion tools were significantly lower for underperforming trainees, but the DOPS and mini-peer assessment tool could not discriminate between trainees of different competency.

Prior to this thesis, only one study has evaluated the DOPS for a specific procedure. Barton et al designed a task-specific tool for colonoscopy procedures, modifying the original pilot General Medical Council DOPS form through a Delphi process and expert consensus.[157]

However, the candidates being assessed were not trainees, but rather experienced experts: the inclusion criteria required the endoscopist to have > 500 previous colonoscopies, > 90% success in reaching the caecum, with sufficient speed and smoothness of procedural skill that minimal sedation was required for their patients (< 5mg midazolam and < 50mg pethidine). In the demographics, candidates had a mean of 2873 previous procedures, and one candidate was excluded having done > 25,000 colonoscopies. This uniformity of skill level limits the usefulness and reliability results of this study.

In anaesthesia, a DOPS tool was incorporated into the assessment schedule of the UK anaesthesia training program, but as of late 2015 a new DOPS form is being introduced. This new form has simplified the scoring system, removed the original summative global satisfactory/unsatisfactory item, and added more opportunities for written feedback, with a stated intention to re-orientate the tool for formative assessment only.[232]

In the ANZCA curriculum, a generic DOPS tool was introduced in 2013 for the assessment of all procedural skills, including UGRA.[119] In two studies of this thesis, psychometric evaluation was performed on the ANZCA DOPS when used to assess UGRA procedures.

1.7.8 Mini-clinical evaluation exercise

The American Board of Internal Medicine (ABIM) introduced the clinical evaluation exercise (CEX) in 1972 as a 'long case' assessment tool. This aimed to measure a trainee's performance in taking a thorough history and examination of a real patient, presentation skills, and articulating the management plan to a single examiner. However, this tool suffered from poor feasibility (2 hours per episode) and poor reliability of 0.0 to 0.61 (biases from case- and content-specificity, and poor inter-rater consistency).[233, 234]

In 1995, Norcini et al redesigned the tool as the mini-CEX, a 10-20 minute version with an emphasis on prioritising diagnosis and management, requiring a focussed history and examination.[235] Trainees are expected to perform several mini-CEX with different assessors. Similar to the GRS and DOPS, items address both technical and non-technical skills, and are scored using Likert scales. Unlike DOPS, the mini-CEX has been more systematically evaluated[236, 237] and has been used in multiple specialties including cardiology,[238] ophthalmology,[239] internal medicine,[240] and recently in anaesthesia.[159]

Weller et al studied the use of the mini-CEX as a WBA in the operating theatre, labour ward, pain round, and pre-anaesthesia clinic.[159] 38 anaesthesia trainees undertook 297 mini-CEX assessments, scored by 58 assessors. Generalisability theory was used to evaluate reliability in this study. The authors found that assessor stringency (doves, with a scale limitation bias towards high scores; and hawks, with scale limitation towards low scores) contributed 40% of error variance. Equally, case-specificity contributed another 40% of variance, while assessor subjective bias (halo effect, where the general impression of the trainee effects the actual scoring for items) contributed 15%.

In these studies, there was good support for the use of the mini-CEX as a formative assessment tool. The domains encompass multiple areas, making this tool conducive for structured feedback. There is evidence for adequate feasibility, though the initial UK General Medical Council pilot noted that the mini-CEX was the least practical compared to the MSF and DOPS tools.[156] If used for summative assessment, the concern is that reliability is modest. To improve reliability, authors have recommended a range of 6 to 16 mini-CEX episodes for each trainee to compensate for variances in scoring.[118, 156]

There are currently no studies using the mini-CEX for regional anaesthesia skills assessment.

1.7.9 Non-technical skills in anaesthesia

Anaesthesia has been noted to share with other occupations such as aviation, air traffic control, and the nuclear power generation industry, a safety-focussed culture. The human factors, prioritising problem solving, planning, communication, flexibility, and resource management priorities in anaesthesia have been detailed in Fletcher et al's review.[57]

The following year, these authors published the Anaesthesia Non-Technical Skills (ANTS) tool which used a 2-staged GRS to assess 4 categories of leadership, situational awareness, decision making, and team work.[241] Assessors scored 16 individual items, then re-scored the 4 categories holistically and in addition an overall performance item. Concurrently, the Non-Technical Skills for Surgeons (NOTSS) was developed[242] and introduced into the Royal Australasian College of Surgeons and Royal College of Surgeons, United Kingdom.

The major limitation to the use of ANTS is lack of assessor training, which require considerable time and financial investment. A one day training course was found to be insufficient in training assessors to a reliable level.[124] A minimum of 2 days has been recommended, with a known issue being achieving consensus in anaesthesia assessors to reduce inter-rater inconsistency in interpreting and scoring several of the behavioural items.[243, 244] Currently, ANTS has not been incorporated into a formal assessment portfolio in anaesthesia training, although elements of non-technical skill is assessed in existing WBA tools.

1.8 Aims and Objectives of Thesis

The literature review identified many areas in which evidence was lacking in education and training in regional anaesthesia, and specifically for ultrasound-guided regional anaesthesia. Implementation of competency-based anaesthesia curricula has involved the introduction of several assessment tools, for which psychometric evaluation has not been adequately performed. Validation of these tools have been consistently noted as a critical missing step in medical education, not just to ensure appropriate trainee assessment but also to provide researchers with reliable tools to measure the effects of educational interventions.

Evidence does exist for task deconstruction and fractionation, part task training on bench top models, deliberate feedback from expert faculty, structured feedback, briefing and debriefing, to create pre-trained regional anaesthetists.

The initial aim of this research program was to fill gaps in the existing literature, by critically evaluating the current assessment tools used for UGRA performance. The results of the initial studies performed in this doctorate provided some answers and generated an aim to develop a validated assessment tool designed for all regional anaesthesia procedures (Studies One to Four).

Furthering the evidence on training, Study Five sought to determine if different bench top models provided an advantage in learning UGRA needle skills. Study Six added to the limited evidence on the role of psychometric testing to evaluate the innate ability of novices to learn UGRA skills.

In this section, the six studies of this thesis are introduced, a brief rationale and description of methods, and the journal of publication. Each study is then reproduced in full from the journal in Chapters 2 to 7.

1.8.1 Evaluation of the Australian and New Zealand College of Anaesthetists Direct Observation of Procedural Skills tool to assess ultrasound-guided regional anaesthesia performance

In 2013, the ANZCA introduced a competency-based anaesthesia curriculum. Regional anaesthesia was to be assessed using the DOPS tool. This tool had not been evaluated for its psychometric properties for use in UGRA, or within the context of Australian or New Zealand clinical anaesthesia practices.

To evaluate this tool, a prospective, observational study was performed to measure reliability, construct validity, and feasibility of the ANZCA DOPS tool for assessment of UGRA procedures (Chapter 2).[245] This was published as *Psychometric Evaluation of a Direct Observation of Procedural Skills Assessment Tool for Ultrasound-Guided Regional Anaesthesia* in Anaesthesia, 2014.

1.8.2 Inter-assessor reliability of the Australian and New Zealand College of Anaesthetists Direct Observation of Procedural Skills tool to assess ultrasound-guided regional anaesthesia performance

In a follow up study, the ANZCA DOPS was again investigated for reliability. This experiment was set up to determine if a larger group of anaesthetists, who would ordinarily be assessors of trainees in the workplace, had comparable reliability in using DOPS to score trainees performing UGRA procedures.

In this prospective observational study, 49 assessors were recruited to view 2 standardised videos of UGRA performance (Chapter Three).[246] This was published as *Reliability of the Direct Observation of Procedural Skills Assessment Tool for Ultrasound-Guided Regional Anaesthesia* in Anaesthesia and Intensive Care, 2016.

This study was accompanied by two editorials (Appendix One),[247, 248] which placed the study in context of the overall ANZCA curriculum objectives, and discussed the wider significance of WBA tools for summative and formative assessment in postgraduate medical education.

1.8.3 Evaluation of a combined checklist and global rating scale designed for ultrasound-guided regional anaesthesia

In 2012, Cheung et al published a task-specific checklist and GRS designed specifically for assessment of UGRA procedures.[214] This was created using a Delphi process, and did not include initial psychometric analysis or explored its characteristics when used clinically.

This tool was evaluated with a prospective observational trial in 2014, with the primary aims to appraise inter-rater reliability, construct validity, and feasibility (Chapter Four).[215] This study was published as *Evaluation of a Task-specific Checklist and Global Rating Scale for Ultrasound-Guided Regional Anesthesia* in Regional Anesthesia and Pain Medicine, 2014.

A known limitation of the Cheung et al checklist is relevance for UGRA procedures only. The research experiences to date helped inform the creation and evaluation of an assessment tool for all regional anaesthesia procedures, abbreviated as the Regional Anaesthesia Procedural Skills (RAPS) assessment tool.

1.8.4 Design, pilot and validation of the Regional Anaesthesia Procedural Skills (RAPS) assessment tool for all regional anaesthesia procedures

RAPS was intended to be used for UGRA and non-UGRA, peripheral and neuraxial regional anaesthesia techniques. This study described the 3 phases of design, pilot, and clinical validation of the RAPS tool (Chapter Five).[249] This study was published as *Design and Validation of the Regional Anaesthesia Procedural Skills (RAPS) Assessment Tool* in Anaesthesia, 2015.

The methodology of this study was to determine if the RAPS tool exhibited evidence of face and content validity, construct validity, test-retest reliability, external reliability, and feasibility. If proven, RAPS would be a novel assessment tool that can be used for all regional anaesthesia procedures.

Publication of the RAPS paper attracted correspondence, and our reply is included in Appendix Two.[250]

1.8.5 Bench table models for training in ultrasound-guided regional anaesthesia

During the literature review, the role of different *in vitro* model to teach UGRA skills was not clearly defined. Study Five of this research was a randomised controlled trial to determine if different types of bench top models provided an advantage for ultrasound-guided needling skills.[251] This study was published as *A Randomised Controlled Trial Comparing Meat-based with Human Cadaveric Models for Teaching Ultrasound-Guided Regional Anaesthesia* in Anaesthesia, 2016.

The hypothesis for this study was that novices trained on fresh frozen human cadavers, representing the current highest fidelity bench top model, were more proficient and faster in performing UGRA sciatic nerve blocks than novices trained on a low cost, easily accessible pork meat model.

This study employed the following evidence based training principles: deconstructing the UGRA procedure to concentrate on the UGRA needling skill set; deliberate practice by expert faculty with feedback to participants on their individual performance; the feedback was structured and concurrent; faculty and participants were aware of the competency criteria (local anaesthesia deposition above and below the nerve target); and feedback was directed towards participants performing to that endpoint.

This study was accompanied by an editorial (Appendix Three),[252] which discussed our results in the context of models used for ultrasound-guided regional anaesthesia training.

1.8.6 Visuospatial ability and sonography performance relevant for ultrasound-guided regional anaesthesia

It is known that visuospatial ability confers an advantage to performance in other medical skills and procedures. However, the literature review revealed limited evidence to help define the role of psychometric testing in UGRA. The other limitation is that the existing studies used multiple heterogeneous tests leading to difficulty when comparing results.

Study Six was designed to investigate if the innate visuospatial ability of novices influenced the proficiency and efficiency of ultrasonography relevant to UGRA.[253] This study was published as *Visuospatial Ability and Novice Brachial Plexus Sonography Performance* in Acta Anaesthesiologica Scandinavica, 2016.

Importantly, we decided to employ well known tests that have strong population data in our test battery. Standardising the testing regime will help future studies as it defines which cognitive factors are important for UGRA and allows uniformity of assessment.

Having tools that educators can use to identify novices who might struggle with UGRA early in training is helpful. Educators can then more efficiently allocate scarce resources and training strategies to trainees who would benefit most

Chapter 2

Study 1: Evaluation of the ANZCA DOPS tool to assess regional anaesthesia performance

Pages 97-105 of this thesis have been removed as they contain published material. Please refer to the following citation for details of the article contained in these pages.

Watson M. J., Wong D. M., Kluger R., Chuan A., Herrick M. D., Ng I., Castanelli D. J., Lin L., Lansdown A., Barrington M. J. (2014) Psychometric evaluation of a direct observation of procedural skills assessment tool for ultrasound-guided regional anaesthesia. *Anaesthesia* 69 p.604-12.

DOI: 10.1111/anae.12625

Chapter 3

Study 2: Inter-assessor reliability of the ANZCA DOPS tool to assess regional anaesthesia performance

Pages 107-116 of this thesis have been removed as they contain published material. Please refer to the following citation for details of the article contained in these pages.

Chuan A., Thillainathan S., Graham P. L., Jolly B., Wong D. M., Smith N. A., Barrington M. J. (2016) Reliability of the Direct Observation of Procedural Skills assessment tool for ultrasound-guided regional anaesthesia. *Anaesthesia and Intensive Care* 44(2):p. 201-209.

DOI: 10.1177/0310057X1604400206

Chapter 4

Study 3: Evaluation of a combined checklist and global rating scale for ultrasound-guided regional anaesthesia

Pages 117-126 of this thesis have been removed as they contain published material.
Please refer to the following citation for details of the article contained in these pages.

Wong D. M., Watson M. J., Kluger R., Chuan A., Herrick M. D., Ng I., Castanelli D. J., Lin L., Lansdown A., Barrington M. J. (2014) Evaluation of a task-specific checklist and global rating scale for ultrasound-guided regional anesthesia. *Regional Anesthesia and Pain Medicine.* 39(5) p.399-408.

DOI: 10.1097/AAP.0000000000000126

Chapter 5

Study 4: Design, pilot and validation of the Regional Anaesthesia Procedural Skills (RAPS) assessment tool

Pages 128-138 of this thesis have been removed as they contain published material. Please refer to the following citation for details of the article contained in these pages.

Chuan A., Graham P. L., Wong D. M., Barrington M. J., Auyong D. B., Cameron A. J. D., Lim Y. C., Pope L., Germanoska B., Forrest K., Royse C. F. (2015) Design and validation of the Regional Anaesthesia Procedural Skills Assessment Tool. *Anaesthesia* 70(12) p.1401-1411.

DOI: 10.1111/anae.13266

Chapter 6

Study 5: Bench table models for training in ultrasound-guided regional anaesthesia

Pages 140-148 of this thesis have been removed as they contain published material. Please refer to the following citation for details of the article contained in these pages.

Chuan A., Lim Y. C., Aneja H., Duce D. A., Appleyard R., Forrest K., Royse C. F. (2016) A randomised controlled trial comparing meat-based with human cadaveric models for teaching ultrasound-guided regional anaesthesia. *Anaesthesia* 71(8) p.921-929

DOI: 10.1111/anae.13446

Chapter 7

Study 6: Visuospatial ability and sonography performance relevant for ultrasound-guided regional anaesthesia

Pages 150-158 of this thesis have been removed as they contain published material. Please refer to the following citation for details of the article contained in these pages.

Duce N. A., Gillett L., Descallar J., Tran M. T., Siu S. C.M., Chuan A. (2016) Visuospatial ability and novice brachial plexus sonography performance. *Acta Anaesthesiologica Scandinavica* 60(8) p.1161-1169

DOI: 10.1111/aas.12757

Chapter 8

Summary and Conclusion

8.1 Summary of the literature review

Regional anaesthesia has changed greatly over history. There has always been a close interaction between innovative ideas, introduction of novel drugs, technological improvements, and new equipment that marked evolutionary changes in regional anaesthesia practice. In the last two decades, the technological advance has been the introduction of ultrasound.

This change has taken place concurrently with the transition in postgraduate medical education towards competency-based curricula. As a consequence, competency in UGRA procedures is viewed as the satisfactory performance of technical and non-technical skill sets. These skill sets have been defined in consensus documents of RA societies, and incorporated into national frameworks describing training in anaesthesia.

These skill sets were already published at the commencement of this research program, but there were different levels of evidence on how best to deliver the education and training required to attain these competencies.

No single medical education theory can comprehensively explain the process of learning technical and non-technical skills by trainees. Relevant theories provide insights in understanding different stages in learning, and helps trainers design suitable programs and assessments that test disparate knowledge, procedural and behavioural skill sets. In technical skills acquisition, Ericsson observed that expert performance is best achieved through deliberate practice with expert feedback on deconstructed tasks.

However, the literature review showed inconsistencies in studies measuring procedural performance. Due to the relatively recent introduction of UGRA, few tools have been published that assessed UGRA performance, and were heterogeneous in design and methodology. This precluded any standardisation of testing and comparison of results.

The aims of this research program were broad. To even begin comparing UGRA performance to competencies, we required tools that were reliable and validated. This was the basis of several studies evaluating the psychometric properties of existing assessment tools. This led to the design and evaluation of the new RAPS assessment tool. The remaining studies examined specific aspects of education and training in UGRA, such as the utility of psychometric testing, the usefulness of fresh frozen cadaveric models, and the postoperative recovery outcomes of regional anaesthesia.

8.2 Summary of the research findings and relevance to the literature

8.2.1 *Psychometric Evaluation of a Direct Observation of Procedural Skills Assessment Tool for Ultrasound-Guided Regional Anaesthesia*

The aim of this study was to perform a psychometric evaluation of the reliability, validity, feasibility of the direct observation of procedural skills (DOPS) assessment tool when used to assess UGRA procedures. DOPS was introduced in 2013 as part of the new competency-based training curriculum of the Australian and New Zealand College of Anaesthetists. DOPS is to be used to assess all procedural skills, including UGRA blocks. However, the ANZCA DOPS has not been previously validated for its intended purpose.

This study recruited 30 videos of ANZCA trainees performing UGRA clinical procedures on patients. Off line analysis of these videos were performed by 6 experts and known educators in UGRA techniques. Experts were trained in the use of DOPS in pilot assessments. Statistical analysis used ICC(2A,1) to evaluate inter-rater reliability.

This study showed that internal consistency is good, and there was evidence for construct validity. It is feasible to use in a clinical environment, being quick to score. However, external reliability is poor, even despite having trained expert assessors. ICC ranged from 0.10 to 0.49 for both individual items and overall scores, and falls short of the thresholds for either summative or formative assessment tools.

This publication presents a format for psychometric evaluation of workplace-based assessment tools. The study was a crossed design where all assessors viewed the same procedure. The basis of construct validity was on experience and seniority levels of trainees. Classical test theory and appropriate statistical analysis were used for reliability estimates. This first validation study of the ANZCA DOPS showed that there was a potential limitation of low reliability due to assessor variance.

8.2.2 *Reliability of the Direct Observation of Procedural Skills Assessment Tool for Ultrasound-Guided Regional Anaesthesia.*

Validation of an assessment tool is a continuous process, and is based on the overall weight of evidence accrued from multiple studies performed over time and in different contexts. The aim of this study was to re-test the reliability of the ANZCA DOPS tool when used to assess UGRA procedures.

This was achieved by changing the following circumstances. Two scripted videos of UGRA procedures were filmed with actors playing the roles of trainee and patient. The assessors in this study were drawn from anaesthetists attending education sessions on how to score using the ANZCA DOPS, supervisors of training, and departmental WBA trainers. These anaesthetists would ordinarily be whom trainees would approach to complete workplace-based assessments.

49 assessors were recruited to score the 2 videos using the ANZCA DOPS. Using the same ICC(2A,1) analysis, the results showed that reliability ranged from 0.15 to 0.43. The range of coefficients was similar to our first study. We noted that the format of assessment was critical; if 2 assessors were used to simultaneously score the same trainee (rather than a single assessor), their averaged scores would be sufficiently reliable for high stakes assessment. This is again due to assessor variance. Averaging scores would mitigate differences in any one assessor's internal bias or stringency in scoring.

The results of this study has important implications for the use and interpretation of DOPS scores in the ANZCA curriculum. If the primary role of DOPS is for formative assessment, then modification of the tool may be necessary: removal of the pass/fail or overall satisfactory/unsatisfactory questions, or the addition of more prompts for structured feedback. There is precedent, as the UK Royal College of Anaesthetists have introduced a modified DOPS form to orientate their tool towards concentrating on formative feedback.

8.2.3 *Evaluation of a Task-specific Checklist and Global Rating Scale for Ultrasound-Guided Regional Anesthesia*

The aim of this study was to perform a psychometric evaluation of the reliability, validity, feasibility of a combined checklist and GRS assessment tool designed specifically for UGRA procedures. Both components were constructed using a Delphi technique amongst UGRA experts drawn from multiple institutions. The checklist was an original design, while the GRS was modified from an existing tool previously used in a UGRA study. However, at the time of this research program, this checklist and GRS tool had not been validated.

During pilot testing, we found that the original 22-items of the checklist required modification to allow usability by assessors in a clinical context. These modifications included the addition of richer text descriptors to the original checklist items, and defining quality compromising behaviours so that dichotomous scoring is easier (presence or absence of the behaviour). For certain items, these were further deconstructed into smaller component tasks to allow more discrete assessment of procedural performance. For example, the original item "scanning of anatomy and proper identification of target" was deconstructed into 4 smaller distinct steps. This created a 31-item checklist for the purposes of the study.

Methodology and statistical analysis were similar to our other studies. The results showed that this UGRA-specific checklist and GRS was feasible to use in clinical practice, and demonstrated construct validity. Reliability when evaluated with ICC was variable, and below what would be acceptable for assessment: 0.44 for the total score, 0.46 for the checklist score, and 0.27 for the GRS score. Cronbach's α was used for analysis of the individual items, with 8 out of the 31 items reaching > 0.75 for reliability between assessors.

Unlike the ANZCA DOPS form, this tool used a checklist with GRS. Both components scored equally with respect to reliability. The advantage of the itemised checklist, by deconstructing a complex UGRA task into discrete skill sets, helps identify specific weaknesses and strengths. This should allow better structured feedback when the tool is used as a prompt for formative assessment.

8.2.4 *Design and Validation of the Regional Anaesthesia Procedural Skills (RAPS) Assessment Tool.*

Building upon the techniques utilised in the previous 3 studies, RAPS was conceived as a regional anaesthesia assessment tool. It was designed for UGRA, non-US, neuraxial, and peripheral nerve blocks. The aim of this study was to design and validate this new tool, incorporating three phases of construction, pilot, and psychometric evaluation. Each phase would be grounded in a robust methodology and statistical analysis.

In the first phase, a Delphi process was employed to construct a checklist. Expert regional anaesthetists were recruited to undergo iterative rounds of design and consensus to reach a final list of items. To ensure the items were criterion-based, the experts were instructed and provided with ANZCA professional documents relevant for regional anaesthesia performance. These documents describe the knowledge base, skill sets, and behaviours expected of consultant anaesthetists in Australia and New Zealand. These documents were supplemented with consensus guidelines or recommendations from other regional anaesthesia societies such as ASRA and ESRA. This checklist was paired with the existing GRS from previous studies.

In the second phase, the RAPS tool was piloted by 10 experts not involved in the first phase. In a test-retest score reliability method, the expert panel scored 3 videos over 2 time periods. Cohen's κ was used to measure inter-rater reliability in this phase, demonstrating sufficient *a priori* reliability of > 0.70.

In the third phase, 70 videos of clinical block performance was collected and scored by 3 regional anaesthesia experts not involved in the first or second phase. To improve generalisability of RAPS, 2 of the 3 assessors were recruited outside of Australia and New Zealand. ICC was used for evaluation of external reliability.

RAPS demonstrated face validity, content validity, construct validity, test-retest reliability, moderate inter-rater reliability (≥ 0.80), and feasibility. The checklist is divided into logical sections related to task or procedure-specific steps. This flexibility allows removal of sections not necessary for the procedure (eg. ultrasound items when only using surface landmark approaches), while providing a framework for structured feedback. RAPS potentially is useful in clinical assessment of performance, and in education research to measure change after a training intervention.

8.2.5 *A Randomised Controlled Trial Comparing Meat-based with Human Cadaveric Models for Teaching Ultrasound-Guided Regional Anaesthesia.*

The aim of this randomised controlled study was to determine if training on high fidelity cadavers improved performance in UGRA needling skills, versus a low fidelity model constructed from porcine meat and bovine tendon.

57 novices were recruited and allocated to either the cadaver or the porcine model for training. All novices received deliberate practice and structured, individualised feedback from the expert trainer faculty.

Videos were taken of pre- and post-training attempts by novices performing a UGRA sciatic nerve block in a human cadaver. Off line analysis of these videos were scored by 2 blinded assessors. The results showed that neither primary outcome (time taken) or secondary outcomes (errors and sonography image quality) were influenced by the training model.

This implies that a simpler, easily constructed, cheaper model is as effective as a fresh frozen cadaver model to train novices. This may be due to 2 factors: in the early learning phase of UGRA technical skills, the type of model may not be as relevant to gain the necessary motor and cognitive skills. With more clinical experience or exposure, a cadaver model has qualitative advantages such as face validity, observing spread of injectate in specific planes, ergonomics, and practice of different anatomical approaches to the nerve. Secondly, this study adds further evidence that the quality of feedback is more important than the fidelity of the model used for training.

8.2.6 *Visuospatial Ability and Novice Brachial Plexus Sonography Performance.*

This study asked the question of what visuospatial factors influence the performance of sonography relevant to UGRA. We approached this question from the basis of the original description of visuospatial ability in the Cattell-Horn-Carroll cognitive abilities taxonomic system. This helped identify 4 possible narrow stratum visuospatial factors that possibly influenced sonography performance: spatial visualisation, flexibility of closure, spatial relations, and speed of closure. This in turn allowed selection of appropriate, standardised psychometric tests of these 4 factors.

33 sonography novices were recruited to perform brachial plexus sonography on a human model, both before and after a short period of discovery learning. 2 blinded assessors evaluated the videos taken of novices' performances. Each video was scored on scanning technique, quality of scanning, and image quality. Time taken to perform the task was recorded.

This study showed that novices with intermediate and higher visuospatial ability have better proficiency and efficiency in performing sonography. Three visuospatial factors appear to be influential in sonography: spatial visualisation, spatial relations, and speed of closure.

This study provides evidence that identification of novices who might be expected to struggle with the sonography task is possible using psychometric tests. From studies on visuospatial ability and other medical skills, we recognise that experts adopt cognitive constructs that help negate disadvantages from innate "giftedness" in procedural tasks. Thus, rather than exclude low ability trainees from UGRA, identifying these trainees potentially allows efficient allocation of training and teaching. Secondly, the use of a standardised visuospatial test battery allows direct comparisons of different studies, including comparisons to studies evaluating different procedural skills in anaesthesia.

8.3 Strengths and limitations of the research

The studies on assessment tools were designed primarily to evaluate the psychometric properties of the tools. Attempts were made to design the studies for methodological and statistical robustness, as these were known limitations. Assessors and experts were recruited from different institutions and different countries, which adds validity to the study results. Iterative rounds of piloting and consensus scoring amongst the assessors was performed to provide some consistency of definitions. However, there is potential that more extensive training might have benefit with higher consistency in scoring.

Statistical methods for reliability were based on classical test theory. In particular, the studies used ICC as the primary statistic for external reliability. The test results are easily interpreted, but given the multiple forms of the statistic, care was taken to apply the most applicable type to the study data. Generalisability theory would have provided an alternative, potentially richer results allowing interpretation of different error variances in reliability. However, there was difficulty in finding appropriate statistical expertise, as there is a lack of familiarity with this complex analysis.

Assessment tools require ongoing evidence that these tools are still valid for the purposes that they were intended. As such, the studies represent initial evaluations of tools when used for the purpose of UGRA procedural assessment. The RAPS tool has been shown to have good psychometric properties and moderate reliability, and has potential to fulfil requirements for clinical assessment of RA procedures as well as a research tool. However, further studies using RAPS is required to adequately understand this tool, with larger trainee and assessor samples in different institutions.

The study on bench top models shows that new trainees or novices to UGRA can be taught needle visibility and tracking skills using an easily sourced and inexpensive model. The training effect of deliberate practice and feedback from an expert faculty cannot be excluded. On average, 30 attempts are required for novices to be accurate in depositing injectate above and below a nerve in a bench top model. Despite this, several limitations exist: this was a controlled environment with a prepared cadaver, this was a single task of the overall UGRA procedure, and time was the primary outcome.

When examining the error rates and types, novices were still committing a median of 16 errors including advancing without needle visualisation, loss of needle visibility, and multiple needle passes. These errors have been noted previously, but has not been scrutinised by studies with the primary aim of correcting these quality compromising behaviours.

The study on visuospatial skills allowed identification of innate factors which influences proficiency in sonography skills. The four visuospatial factors chosen to be tested was determined *a priori*, based on review of the literature for other medical procedural tasks similar to UGRA. The choice of tests was deliberate and will allow comparisons with population norms and across different studies.

There were three limitations: psychomotor ability was not assessed, the task was specifically on sonography rather than UGRA needle skills, and there was no educational intervention for the low ability novices. Nonetheless the study opens another avenue of research into the factors that contribute to UGRA learning.

Future studies should aim to combine visuospatial with psychomotor testing to create a complete psychomotor profile. It also remains to be seen which educational strategy is best suited to help low ability novices gain the cognitive schema that helps with understanding and interpreting ultrasound imagery.

8.4 Summary: Future of training in UGRA

Regional anaesthesia, and UGRA in particular, is a difficult procedural task that draws upon large and varied skill sets. For example, volume of practice studies suggests that even after 60 UGRA blocks there is a large variability in performance and error rates. RA techniques are also the hardest procedural skill for anaesthetists to learn, compared to other skills such as arterial cannulation, central venous catheterisation, and laryngoscopy and tracheal intubation.

The existence of the learning curve dictates that the goal of achieving competency requires time to practice. There is evidence that this learning can occur safely, by practicing *in vitro* deconstructed tasks on bench top models prior to clinical contact. There is much interest on the role of simulation, and the transfer of skills from bench top to clinical practice.

It is postulated that simulation creates a learner-orientated environment, allowing repetition, a graduated exposure from simple partial tasks to complex integrated procedures, and opportunities for feedback and reflection. Simulation learning is also more efficient than volume of caseload, given workplace changes are reducing clinical exposure compared to previous generations of trainees.

Another exciting possibility of research is individualising training for each trainee. Individuals learn at different rates and are receptive to different educational formats. Currently, training is delivered as a one-size-fits-all approach. Identifying novices using psychometric tests is one such attempt to determine which trainees might benefit from specialised educational intervention. Validated assessment tools assist in identifying weaknesses and strengths to allow for individualised feedback. To deliver such high quality and flexible teaching, trainers themselves have to be trained, which has been identified as an ongoing weakness in medical education.

In this context, research into education and training in UGRA is very much in infancy. This doctorate has assessed the assessment tools that are used to measure UGRA performance, designed a new assessment tool for regional anaesthesia, and has added evidence for partial task trainers and for the role of psychometric testing. However, there are still many unresolved questions, and many avenues of inquiry. Rather than being daunted, these represent great opportunities for collaborations and further advancement of the specialty.

Appendices

Appendix One

Editorials accompanying Study Two (chapter 3) in *Anaesthesia and Intensive Care, 2016.*

Editorial One: Assessing anaesthesia trainees at work: opportunities and challenges

Editorial Two: Reliability of numerical scales used for Direct Observation of Procedural Skills

Appendix Two

correspondence and reply to Study Four (chapter 5) in *Anaesthesia, 2016*

Regional anaesthesia competency assessment

Regional anaesthesia competency assessment - a reply

Appendix Three

Editorial accompanying Study Five (chapter 6) in *Anaesthesia, 2016*.

Cadaveric Training – The Solution for Ultrasound-Guided Regional Anaesthesia?

Appendix Four

Ethics approval letters for studies

Pages 171-185 of this thesis have been removed as they contain published material. Please refer to the following citation for details of the articles contained in these pages.

Roberts, L. J., Jones, O. (2016). Assessing anaesthesia trainees at work: opportunities and challenges. *Anaesthesia and Intensive Care*, 44(2), p. 194-197.
doi: 10.1177/0310057X1604400204

Garden, A. L. (2016).Regionalability of numerical scales used for direct observation of procedural skills. *Anaesthesia and Intensive Care,* 44(2), p. 198-200.
doi: 10.1177/0310054X1604400205

Ahmed, O., O'Donnell, B., Shorten, G., Gallagher, A. (2016) Regional anaesthesia competency assessment. *Anaesthesia,* 71(4), p. 472-473.
doi:10.1111/anae.13393

Chuan, A., Graham, P. L., Forrest, K., Barrington, M. J., Royse, C. F., Wong, D. M., Cameron, A. J. D., Lim, Y. C., Auyong, D. B. (2016) Regional anaesthesia competency assessment - a reply. Anaesthesia, 71(4), p. 473-474.
doi:10.1111/anae.13430

Gupta, A. K., Morton, J. R. (2016) Cadaveric training - the solution for ultrasound-guided regional anaesthesia? *Anaesthesia,* 71(8), p. 874-878.
doi: 10.1111/anae.13538

Appendix Four of this thesis has been removed as it may contain sensitive/confidential content

References

1 Goerig M, Bacon D, van Zundert A. Carl Koller, Cocaine, and Local Anesthesia: Some Less Known and Forgotten Facts. *Reg Anesth Pain Med* 2012; **37**: 318-24

2 Koller C. On the use of cocaine for producing anaesthesia on the eye. *Lancet*; **124**: 990-2

3 Wood A. New method of treating neuralgia by the direct application of opiates to the painful points. *Edinburgh Medical and Surgical Journal* 1855; **82**: 265-81

4 Borgeat A. All Roads Do Not Lead to Rome. *Anesthesiology* 2006; **105**: 1-2

5 Denny N, Harrop-Griffiths W. Location, location, location! Ultrasound imaging in regional anaesthesia. *Br J Anaesth* 2005; **94**: 1-3

6 Cushing H. I. On the Avoidance of Shock in Major Amputations by Cocainization of Large Nerve-Trunks Preliminary to their Division. With Observations on Blood-Pressure Changes in Surgical Cases. *Ann Surg* 1902; **36**: 321-45

7 Kulenkampff D. Brachial plexus anaesthesia: Its indications, technique, and dangers. *Ann Surg* 1928; **87**: 883-91

8 Goerig M, Agarwal K. Georg Perthes--the man behind the technique of nerve-tracer technology. *Reg Anesth Pain Med* 2000; **25**: 296-301

9 Marhofer P, Harrop-Griffiths W. Nerve location in regional anaesthesia: finding what lies beneath the skin. *Br J Anaesth* 2011; **106**: 3-5

10 Greenblatt G, Denson J. Needle nerve stimulatorlocator: nerve blocks with a new instrument for locating nerves. *Anesth Analg* 1962; **41**: 599-602

11 Pither C, Raj P, Ford D. The Use of Peripheral Nerve Stimulators for Regional Anesthesia A Review of Experimental Characteristics, Technique, and Clinical Applications. *Reg Anesth* 1985; **10**: 49-58

12 Selander D, Edshage S, Wolff T. Paresthesiae or no paresthesiae? Nerve lesions after axillary blocks. *Acta Anaesth Scand* 1979; **23**: 27-33

13 Klein S, Melton M, Grill W, Nielsen K. Peripheral nerve stimulation in regional anesthesia. *Reg Anesth Pain Med* 2012; **37**: 383-92

14 VadeBoncouer T, Riegler F. In defense of the nerve stimulator. *Reg Anesth Pain Med* 1998; **23**: 229-30

15 Benhamou D. Axillary Plexus Block Using Multiple Nerve Stimulation: A European View. *Reg Anesth Pain Med* 2001; **26**: 495-8

16 Lang S. The art and science of using a peripheral nerve stimulator: how close is close enough? *Reg Anesth Pain Med* 2002; **27**: 330-2

17 Hadzic A, Vloka J, Hadzic N, Thys D, Santos A. Nerve stimulators used for peripheral nerve blocks vary in their electrical characteristics. *J Am Soc Anesthesiol* 2003; **98**: 969-74

18 Sites B, Gallagher J, Sparks M. Ultrasound-guided popliteal block demonstrates an atypical motor response to nerve stimulation in 2 patients with diabetes mellitus. *Reg Anesth Pain Med* 2003; **28**: 479-82
19 Keyl C, Held T, Albiez G, Schmack A, Wiesenack C. Increased electrical nerve stimulation threshold of the sciatic nerve in patients with diabetic foot gangrene: A prospective parallel cohort study. *Eur J Anaesth* 2013; **30**: 435-40
20 Choyce A, Chan V, Middleton W, Knight P, Peng P, McCartney C. What Is the Relationship Between Paresthesia and Nerve Stimulation for Axillary Brachial Plexus Block? *Reg Anesth Pain Med* 2001; **26**: 100-4
21 Neal J. How Close Is Close Enough? Defining the "Paresthesia Chad". *Reg Anesth Pain Med* 2001; **26**: 97-9
22 Al-Nasser B, Hubert C, Negre M. Role of local anesthetic spread pattern and electrical stimulation in ultrasound-guided musculocutaneous nerve block. *J Clin Anesth* 2010; **22**: 334-9
23 Perlas A, Niazi A, McCartney C, Chan V, Xu D, Abbas S. The sensitivity of motor response to nerve stimulation and paresthesia for nerve localization as evaluated by ultrasound. *Reg Anesth Pain Med* 2006; **31**: 445-50
24 Urmey W, Stanton J. Inability to Consistently Elicit a Motor Response following Sensory Paresthesia during Interscalene Block Administration. *Anesthesiology* 2002; **96**: 552-4
25 Fredrickson M. The sensitivity of motor response to needle nerve stimulation during ultrasound guided interscalene catheter placement. *Reg Anesth Pain Med* 2008; **33**: 291-6
26 La Grange P, Foster P, Pretorius L. Application of the Doppler ultrasound bloodflow detector in supraclavicular brachial plexus block. *Br J Anaesth* 1978; **50**: 965-7
27 PL T, V S. Ultrasonographic study of the spread of local anaesthetic during axillary brachial plexus block. *Br J Anaesth* 1989; **63**: 326-9
28 Kapral S, Krafft P, Eibenberger K, Fitzgerald R, Gosch M, Weinstabl C. Ultrasound-Guided Supraclavicular Approach for Regional Anesthesia of the Brachial Plexus. *Anesth Analg* 1994; **78**: 507-13
29 Chan V. Applying ultrasound imaging to interscalene brachial plexus block. *Reg Anesth Pain Med* 2003; **28**: 340-3
30 Perlas A, Chan V, Simons M. Brachial plexus examination and localization using ultrasound and electrical stimulation: a volunteer study. *Anesthesiology* 2003; **99**: 429-35
31 Sandhu N, Capan L. Ultrasound-guided infraclavicular brachial plexus block. *Br J Anaesth* 2002; **89**: 254-9
32 Ootaki C, Hayashi H, Amano M. Ultrasound-guided infraclavicular brachial plexus block: an alternative technique to anatomical landmark-guided approaches. *Reg Anesth Pain Med* 2000; **25**: 600-4

33 Marhofer P, Schrogendorfer K, Koinig H, Kapral S, Weinstabl C, Mayer N. Ultrasonographic Guidance Improves Sensory Block and Onset Time of Three-in-One Blocks. *Anesth Analg* 1997; **85**: 854-7

34 Grau T, Leipold R, Conradi R, Martin E, Motsch J. Ultrasound imaging facilitates localization of the epidural space during combined spinal and epidural anesthesia. *Reg Anesth Pain Med* 2001; **26**: 64-7

35 Chin K, Karmakar M, Peng P. Ultrasonography of the adult thoracic and lumbar spine for central neuraxial blockade. *Anesthesiology* 2011; **114**: 1459-85

36 Halpern S, Banerjee A, Stocche R, Glanc P. The use of ultrasound for lumbar spinous process identification: A pilot study. *Can J Anaesth* 2010; **57**: 817-22

37 Krediet A, Moayeri N, van Geffen G, et al. Different Approaches to Ultrasound-guided Thoracic Paravertebral Block: An Illustrated Review. *Anesthesiology* 2015; **123**: 459-74

38 Hebbard P, Fujiwara Y, Shibata Y, Royse C. Ultrasound-guided transversus abdominis plane (TAP) block. *Anaesth Intens Care* 2007; **35**: 616-7

39 Hebbard P, Barrington M, Vasey C. Ultrasound-guided continuous oblique subcostal transversus abdominis plane blockade: description of anatomy and clinical technique. *Reg Anesth Pain Med* 2010; **35**: 436-41

40 Blanco R. The 'pecs block': a novel technique for providing analgesia after breast surgery. *Anaesthesia* 2011; **66**: 847-8

41 Visoiu M, Yakovleva N. Continuous postoperative analgesia via quadratus lumborum block – an alternative to transversus abdominis plane block. *Pediatr Anesth* 2013; **23**: 959-61

42 McCartney C, Lin L, Shastri U. Evidence basis for the use of ultrasound for upper-extremity blocks. *Reg Anesth Pain Med* 2010; **35**: S10-5

43 Choi S, McCartney C. Evidence Base for the Use of Ultrasound for Upper Extremity Blocks: 2014 Update. *Reg Anesth Pain Med* 2016; **41**: 242-50

44 Salinas F. Ultrasound and review of evidence for lower extremity peripheral nerve blocks. *Reg Anesth Pain Med* 2010; **35**: S16-25

45 Salinas F. Evidence Basis for Ultrasound Guidance for Lower-Extremity Peripheral Nerve Block: Update 2016. *Reg Anesth Pain Med* 2016; **41**: 261-74

46 Abrahams M, Horn J, Noles L, Aziz M. Evidence-Based Medicine: Ultrasound Guidance for Truncal Blocks. *Reg Anesth Pain Med* 2010; **35**: S36-S42

47 Abrahams M, Derby R, Horn J. Update on Ultrasound for Truncal Blocks: A Review of the Evidence. *Reg Anesth Pain Med* 2016; **41**: 275-88

48 Perlas A, Chaparro L, Chin K. Lumbar Neuraxial Ultrasound for Spinal and Epidural Anesthesia: A Systematic Review and Meta-Analysis. *Reg Anesth Pain Med* 2016; **41**: 251-60

49 Lewis S, Price A, Walker K, McGrattan K, Smith A. Ultrasound guidance for upper and lower limb blocks. *Cochrane Database of Systematic Reviews* 2015
50 Guay J, Suresh S, Kopp S. The use of ultrasound guidance for perioperative neuraxial and peripheral nerve blocks in children. *Cochrane Database of Systematic Reviews* 2016
51 Hopkins P. Ultrasound guidance as a gold standard in regional anaesthesia. *Br J Anaesth* 2007; **98**: 299-301
52 Sites B, Brull R, Chan V, et al. Artifacts and pitfall errors associated with ultrasound-guided regional anesthesia. Part I: understanding the basic principles of ultrasound physics and machine operations. *Reg Anesth Pain Med* 2007; **32**: 412-8
53 Sites B, Brull R, Chan V, et al. Artifacts and pitfall errors associated with ultrasound-guided regional anesthesia. Part II: a pictorial approach to understanding and avoidance. *Reg Anesth Pain Med* 2007; **32**: 419-33
54 American Institute of Ultrasound in Medicine. *AIUM Practice Parameter for the Performance of Selected Ultrasound-Guided Procedures*. Laurel, Maryland, USA: American Institute of Ultrasound in Medicine, 2014
55 Sites B, Chan V, Neal J, et al. The American Society of Reg Anesth Pain Med and the European Society of Regional Anaesthesia and Pain Therapy joint committee recommendations for education and training in ultrasound-guided regional anesthesia. *Reg Anesth Pain Med* 2010; **35**: S74-80
56 Smith A, Pope C, Goodwin D, Mort M. What defines expertise in regional anaesthesia? An observational analysis of practice. *Br J Anaesth* 2006; **97**: 401-7
57 Fletcher G, McGeorge P, Flin R, Glavin R, Maran N. The role of non-technical skills in anaesthesia: a review of current literature. *Br J Anaesth* 2002; **88**: 418-29
58 Van Gessel E, Mellin-Olsen J, Ostergaard H, Niemi-Murola L, Education, Training Standing Committee EBoARaIC. Postgraduate training in anaesthesiology, pain and intensive care: the new European competence-based guidelines. *Eur J Anaesth* 2012; **29**: 165-8
59 Frank J, Snell L, Sherbino J. *CanMEDS 2015 Physician Competency Framework*. Ottawa: Royal College of Physicians and Surgeons of Canada, 2015
60 *ANZCA Curriculum Framework*. Melbourne, Australia: Australian and New Zealand College of Anaesthetists, 2013
61 Ebert T, Fox C. Competency-based Education in Anesthesiology: History and Challenges. *Anesthesiology* 2014; **120**: 24-31
62 Tetzlaff J. Assessment of Competency in Anesthesiology. *Anesthesiology* 2007; **106**: 812-25
63 Marhofer P. Ultrasound Guidance in Regional Anaesthesia: Principles and practical implementation. Oxford, UK: Oxford University Press, 2010

64 Brull R, Macfarlane A, Tse C. Practical Knobology for Ultrasound-Guided Regional Anesthesia. *Reg Anesth Pain Med* 2010; **35**: S68-S73

65 Sites B, Macfarlane A, Sites V, et al. Clinical Sonopathology for the Regional Anesthesiologist: Part 1: Vascular and Neural. *Reg Anesth Pain Med* 2010; **35**: 272-80

66 Sites B, Macfarlane A, Sites V, et al. Clinical Sonopathology for the Regional Anesthesiologist: Part 2: Bone, Viscera, Subcutaneous Tissue, and Foreign Bodies. *Reg Anesth Pain Med* 2010; **35**: 281-9

67 Lichtenstein D, Menu Y. A bedside ultrasound sign ruling out pneumothorax in the critically iii : Lung sliding. *Chest* 1995; **108**: 1345-8

68 Ihnatsenka B, Boezaart A. Ultrasound: Basic understanding and learning the language. *Int J Should Surg* 2010; **4**: 55-62

69 Chuan A, Scott D. *Regional Anaesthesia. A Pocket Guide*. United Kingdom: Oxford University Press, 2014

70 Chin K, Perlas A, Chan V, Brull R. Needle Visualization in Ultrasound-Guided Regional Anesthesia: Challenges and Solutions. *Reg Anesth Pain Med* 2008; **33**: 532-44

71 Sehmbi H, Perlas A. Basics of Ultrasound Imaging. *Regional Nerve Blocks in Anesthesia and Pain Therapy: Traditional and Ultrasound-Guided Techniques*. Cham: Springer International Publishing, 2015; 27-56

72 *PS37: Guidelines for Health Practitioners Administering Local Anaesthesia*. Melbourne, Australia: Australian and New Zealand College of Anaesthetists, 2013

73 *PS03: Guidelines for the Management of Major Regional Anaesthesia*. Melbourne, Australia: Australian and New Zealand College of Anaesthetists, 2014

74 *PS26: Guidelines on Consent for Anaesthesia or Sedation*. Melbourne, Australia: Australian and New Zealand College of Anaesthetists, 2005

75 *PS08: Statement on the Assistant or the Anaesthetist*. Melbourne, Australia: Australian and New Zealand College of Anaesthetists, 2015

76 Maslow A. *Motivation and Personality*. New York: Harper and Brothers, 1954

77 Forrest K, McKimm J, Edgar S. *Essential Simulation in Clinical Education*. Hoboken, USA: Wiley Blackwell, 2013

78 Tay L, Diener E. Needs and subjective well-being around the world. *J Pers Soc Psychol* 2011; **101**: 354-65

79 Kolb D. *Experiential Learning: Experience as the Source of Learning and Development*. Engelwood Cliffs, New Jersey: Prentice-Hall, 1984

80 Smits P, Verbeek J, Nauta M, Ten Cate T, Metz J, van Dijk F. Factors predictive of successful learning in postgraduate medical education. *Med Educ* 2004; **38**: 758-66

81 Vygotsky L. *Mind in society: The development of higher psychological processes*. Cambridge, MA, USA: Harvard University Press, 1978

82 Wood D, Bruner J, Ross G. The role of tutoring in problem solving. *J Child Psychol Psyc* 1976; **17**: 89-100
83 Slater R, Castanelli D, Barrington M. Learning and teaching motor skills in regional anesthesia: a different perspective. *Reg Anesth Pain Med* 2014; **39**: 230-9
84 Miller G. The assessment of clinical skills/competence/performance. *Acad Med* 1990; **65**: S63-7
85 Australian Medical Council. 2016. Available from http://wbaonline.amc.org.au/about/ (accessed 20 May 2016 2016)
86 Fitts P, Posner M. *Human performance*. Oxford, UK: Brooks/Cole, 1967
87 Taylor J, Ivry R. The role of strategies in motor learning. *Ann NY Acad Sci* 2012; **1251**: 1-12
88 Lohse K, Wadden K, Boyd L, Hodges N. Motor skill acquisition across short and long time scales: a meta-analysis of neuroimaging data. *Neuropsychologia* 2014; **59**: 130-41
89 Patel R, Spreng R, Turner G. Functional brain changes following cognitive and motor skills training: a quantitative meta-analysis. *Neurorehab Neural Re* 2013; **27**: 187-99
90 Konrad C, Schupfer G, Wietlisbach M, Gerber H. Learning Manual Skills in Anesthesiology: Is There a Recommended Number of Cases for Anesthetic Procedures? *Anesth Analg* 1998; **86**: 635-9
91 Dreyfus S, Dreyfus H. A five stage model of the mental activities involved in directed skill acquisition. *California University Berkeley Operations Research Center* 1980
92 Dreyfus S. The Five-Stage Model of Adult Skill Acquisition. *B Sci Technol Soc* 2004; **24**: 177-81
93 Green M, Aagaard E, Caverzagie K, et al. Charting the Road to Competence: Developmental Milestones for Internal Medicine Residency Training. *J Grad Med Ed* 2009; **1**: 5-20
94 *The Anesthesiology Milestone Project*. Accreditation Council for Graduate Medical Education and the American Board of Anesthesiology, 2015
95 Glavin R, Maran N. Editorial I: Development and use of scoring systems for assessment of clinical competence. *Br J Anaesth* 2002; **88**: 329-30
96 Khan K, Ramachandran S. Conceptual framework for performance assessment: competency, competence and performance in the context of assessments in healthcare--deciphering the terminology. *Med Teach* 2012; **34**: 920-8
97 Kirkpatrick K, MacKinnon R. Technology-enhanced learning in anaesthesia and educational theory. *Continuing Education in Anaesthesia, Critical Care & Pain* 2012; **12**: 263-7
98 Aggarwal R, Grantcharov T, Darzi A. Framework for systematic training and assessment of technical skills. *J Am Coll Surgeons* 2007; **204**: 697-705

99 Clark R, Pugh C, Yates K, Inaba K, Green D, Sullivan M. The use of cognitive task analysis to improve instructional descriptions of procedures. *J Surg Res* 2012; **173**: e37-42
100 Canopy E, Evans M, Boehler M, Roberts N, Sanfey H, Mellinger J. Interdisciplinary cognitive task analysis: a strategy to develop a comprehensive endoscopic retrograde cholangiopancreatography protocol for use in fellowship training. *Am J Surg* 2015; **210**: 710-4
101 Velmahos G, Toutouzas K, Sillin L, et al. Cognitive task analysis for teaching technical skills in an inanimate surgical skills laboratory. *Am J Surg* 2004; **187**: 114-9
102 Sullivan M, Ortega A, Wasserberg N, Kaufman H, Nyquist J, Clark R. Assessing the teaching of procedural skills: can cognitive task analysis add to our traditional teaching methods? *Am J Surg* 2008; **195**: 20-3
103 O'Sullivan O, Aboulafia A, Iohom G, O'Donnell B, Shorten G. Proactive Error Analysis of Ultrasound-Guided Axillary Brachial Plexus Block Performance. *Reg Anesth Pain Med* 2011; **36**: 502-7
104 Gallagher A, Ritter E, Champion H, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 2005; **241**: 364-72
105 Ericsson K. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004; **79**: S70-81
106 Palter V, Grantcharov T. Individualized deliberate practice on a virtual reality simulator improves technical performance of surgical novices in the operating room: a randomized controlled trial. *Ann Surg* 2014; **259**: 443-8
107 Marcus H, Vakharia V, Kirkman M, Murphy M, Nandi D. Practice makes perfect? The role of simulation-based deliberate practice and script-based mental rehearsal in the acquisition and maintenance of operative neurosurgical skills. *Neurosurgery* 2013; **72 Suppl 1**: 124-30
108 Hastings R, Rickard T. Deliberate practice for achieving and maintaining expertise in anesthesiology. *Anesth Analg* 2015; **120**: 449-59
109 McGaghie W, Issenberg S, Cohen E, Barsuk J, Wayne D. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad Med* 2011; **86**: 706-11
110 Crochet P, Aggarwal R, Dubb S, et al. Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills. *Ann Surg* 2011; **253**: 1216-22
111 Johnson K, Syroid N, Drews F, et al. Part Task and variable priority training in first-year anesthesia resident education: a combined didactic and simulation-based approach to

improve management of adverse airway and respiratory events. *Anesthesiology* 2008; **108**: 831-40

112 Bruppacher H, Alam S, LeBlanc V, et al. Simulation-based training improves physicians' performance in patient care in high-stakes clinical setting of cardiac surgery. *Anesthesiology* 2010; **112**: 985-92

113 de Oliveira Filho G, Helayel P, da Conceicao D, Garzel I, Pavei P, Ceccon M. Learning curves and mathematical models for interventional ultrasound basic skills. *Anesth Analg* 2008; **106**: 568-73, table of contents

114 Barrington M, Wong D, Slater B, Ivanusic J, Ovens M. Ultrasound-guided regional anesthesia: how much practice do novices require before achieving competency in ultrasound needle visualization using a cadaver model. *Reg Anesth Pain Med* 2012; **37**: 334-9

115 Nicol D, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud High Ed* 2006; **31**: 199-218

116 Roberts N, Williams R, Kim M, Dunnington G. The briefing, intraoperative teaching, debriefing model for teaching in the operating room. *J Am Coll Surgeons* 2009; **208**: 299-303

117 Smith H, Kopp S, Jacob A, Torsher L, Hebl J. Designing and implementing a comprehensive learner-centered regional anesthesia curriculum. *Reg Anesth Pain Med* 2009; **34**: 88-94

118 Norcini J. Current perspectives in assessment: the assessment of performance at work. *Med Educ* 2005; **39**: 880-9

119 *ANZCA Handbook for Training and Accreditation*. Melbourne, Australia: Australian and New Zealand College of Anaesthetists, 2014

120 *Curriculum for a CCT in Anaesthetics*. The Royal College of Anaesthetists, 2010

121 Bould M, Crabtree N, Naik V. Assessment of procedural skills in anaesthesia. *Br J Anaesth* 2009; **103**: 472-83

122 Dannefer E. Beyond assessment of learning toward assessment for learning: Educating tomorrow's physicians. *Med Teach* 2013; **35**: 560-3

123 Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach* 2011; **33**: 206-14

124 Graham J, Hocking G, Giles E. Anaesthesia Non-Technical Skills: can anaesthetists be trained to reliably use this behavioural marker system in 1 day? *Br J Anaesth* 2010; **104**: 440-5

125 Holmboe E, Hawkins R, Huot S. Effects of training in direct observation of medical residents' clinical competence: a randomized trial. *Ann Intern Med* 2004; **140**: 874-81

126 Crossley J, Humphris G, Jolly B. Assessing health professionals. *Med Educ* 2002; **36**: 800-4
127 van der Vleuten C, Schuwirth L. Assessing professional competence: from methods to programmes. *Med Educ* 2005; **39**: 309-17
128 Wood T. Assessment not only drives learning, it may also help learning. *Med Educ* 2009; **43**: 5-6
129 Devine O, Harborne A, McManus I. Assessment at UK medical schools varies substantially in volume, type and intensity and correlates with postgraduate attainment. *BMC Med Educ* 2015; **15**: 146
130 Raupach T, Brown J, Anders S, Hasenfuss G, Harendza S. Summative assessments are more powerful drivers of student learning than resource intensive teaching formats. *BMC Med* 2013; **11**: 1-10
131 Raupach T, Harendza S, Anders S, Schuelper N, Brown J. How can we improve teaching of ECG interpretation skills? Findings from a prospective randomised trial. *J Electrocardiol* 2016; **49**: 7-12
132 Kromann C, Jensen M, Ringsted C. The effect of testing on skills learning. *Med Educ* 2009; **43**: 21-7
133 Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach* 2007; **29**: 855-71
134 Veloski J, Boex J, Grasberger M, Evans A, Wolfson D. Systematic review of the literature on assessment, feedback and physicians' clinical performance*: BEME Guide No. 7. *Med Teach* 2006; **28**: 117-28
135 Issenberg S, McGaghie W, Petrusa E, Gordon D, Scalese R. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005; **27**: 10-28
136 Porte M, Xeroulis G, Reznick R, Dubrowski A. Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. *Am J Surg* 2007; **193**: 105-10
137 Xeroulis G, Park J, Moulton C, Reznick R, Leblanc V, Dubrowski A. Teaching suturing and knot-tying skills to medical students: a randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. *Surgery* 2007; **141**: 442-9
138 Kogan J, Conforti L, Bernabeo E, Durning S, Hauer K, Holmboe E. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Med Educ* 2012; **46**: 201-15
139 Govaerts M, Schuwirth L, Van der Vleuten C, Muijtjens A. Workplace-based assessment: effects of rater expertise. *Adv Health Sci Educ* 2011; **16**: 151-65

140 Govaerts M, Van de Wiel M, Schuwirth L, Van der Vleuten C, Muijtjens A. Workplace-based assessment: raters' performance theories and constructs. *Adv Health Sci Educ* 2013; **18**: 375-96

141 Allen B, Sandberg W. Assessment Tools: Searching for Purpose. *Reg Anesth Pain Med* 2015; **40**: 299-300

142 *Appendix One - ANZCA Guidelines on Assessment*. Melbourne, Australia: Australian and New Zealand College of Anaesthetists, 2014

143 Epstein R. Assessment in medical education. *N Engl J Med* 2007; **356**: 387-96

144 Schuwirth L, Colliver J, Gruppen L, et al. Research in assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach* 2011; **33**: 224-33

145 Van der Vleuten C. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ* 1996; **1**: 41-67

146 Downing S. Validity: on meaningful interpretation of assessment data. *Med Educ* 2003; **37**: 830-7

147 Downing S. Reliability: on the reproducibility of assessment data. *Med Educ* 2004; **38**: 1006-12

148 Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna G. Observational tools for assessment of procedural skills: a systematic review. *Am J Surg* 2011; **202**: 469-80

149 Gallagher A, Ritter E, Satava R. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc* 2003; **17**: 1525-9

150 Kogan J, Holmboe E, Hauer K. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA* 2009; **302**: 1316-26

151 Bravo G, Potvin L. Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. *J Clin Epidemiol* 1991; **44**: 381-90

152 Cicchetti D. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychol Assessment* 1994; **6**: 284-90

153 Bartko J. On various intraclass correlation reliability coefficients. *Psychol Bull* 1976; **83**: 762-5

154 McGraw K, Wong S. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; **1**: 30-46

155 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; **86**: 420-8

156 Wilkinson J, Crossley J, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ* 2008; **42**: 364-73
157 Barton J, Corbett S, van der Vleuten C. The validity and reliability of a Direct Observation of Procedural Skills assessment tool: assessing colonoscopic skills of senior endoscopists. *Gastrointest Endosc* 2012; **75**: 591-7
158 Fehr J, Boulet J, Waldrop W, Snider R, Brockel M, Murray D. Simulation-based Assessment of Pediatric Anesthesia Skills. *J Am Soc Anesth* 2011; **115**: 1308-15
159 Weller J, Jolly B, Misur M, et al. Mini-clinical evaluation exercise in anaesthesia training. *Br J Anaesth* 2009; **102**: 633-41
160 Crossley J, Davies H, Humphris G, Jolly B. Generalisability: a key to unlock professional assessment.[Erratum appears in Med Educ. 2003 Jun;37(6):574]. *Med Educ* 2002; **36**: 972-8
161 Briesch A, Swaminathan H, Welsh M, Chafouleas S. Generalizability theory: a practical guide to study design, implementation, and interpretation. *J School Psychol* 2014; **52**: 13-35
162 Woodworth G, Carney P, Cohen J, et al. Development and Validation of an Assessment of Regional Anesthesia Ultrasound Interpretation Skills. *Reg Anesth Pain Med* 2015; **40**: 306-14
163 Kathirgamanathan A, Woods L. Educational tools in the assessment of trainees in anaesthesia. *Continuing Education in Anaesthesia, Critical Care & Pain* 2011; **11**: 138-42
164 Wragg A, Wade W, Fuller G, Cowan G, Mills P. Assessing the performance of specialist registrars. *Clin Med* 2003; **3**: 131-4
165 van Hove P, Tuijthof G, Verdaasdonk E, Stassen L, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg* 2010; **97**: 972-87
166 Varas J, Achurra P, Leon F, et al. Assessment of central venous catheterization in a simulated model using a motion-tracking device: an experimental validation study. *Ann Surg Innov Res* 2016; **10**: 2
167 Clinkard D, Holden M, Ungi T, et al. The development and validation of hand motion analysis to evaluate competency in central line catheterization. *Acad Emerg Med* 2015; **22**: 212-8
168 Chin K, Tse C, Chan V, Tan J, Lupu C, Hayter M. Hand motion analysis using the imperial college surgical assessment device: validation of a novel and objective performance measure in ultrasound-guided peripheral nerve blockade. *Reg Anesth Pain Med* 2011; **36**: 213-9
169 Naik V, Perlas A, Chandra D, Chung D, Chan V. An assessment tool for brachial plexus regional anesthesia performance: establishing construct validity and reliability. *Reg Anesth Pain Med* 2007; **32**: 41-5

170 Norris A, McCahon R. Cumulative sum (CUSUM) assessment and medical education: a square peg in a round hole. *Anaesthesia* 2011; **66**: 250-4
171 Kestin I. A statistical approach to measuring the competence of anaesthetic trainees at practical procedures. *Br J Anaesth* 1995; **75**: 805-9
172 de Oliveira Filho G. The construction of learning curves for basic skills in anesthetic procedures: an application for the cumulative sum method. *Anesth Analg* 2002; **95**: 411-6
173 McGrew K. CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence* 2009; **37**: 1-10
174 Schneider W, McGrew K. The Cattell-Horn-Carroll model of intelligence. In: Flanagan P, Harrison P, eds. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. New York, USA: Guildford Press, 2012; 99-144
175 Ackerman P. Determinants of Individual Differences During Skill Acquisition: Cognitive Abilities and Information Processing. *J Exp Psychol* 1988; **117**: 288-318
176 Voelkle M, Wittmann W, Ackerman P. Abilities and skill acquisition: A latent growth curve approach. *Learn Individ Differ* 2006; **16**: 303-19
177 Hegarty M, Keehner M, Cohen C, Montello D, Lippa Y. The role of spatial cognition in medicine: Applications for selecting and training professionals. In: Allen GL, ed. *Applied spatial cognition*. Mahwah, NJ: Erlbaum, 2007
178 Enochsson L, Westman B, Ritter E, et al. Objective assessment of visuospatial and psychomotor ability and flow of residents and senior endoscopists in simulated gastroscopy. *Surg Endosc* 2006; **20**: 895-9
179 Hedman L, Klingberg T, Enochsson L, Kjellin A, Fellander-Tsai L. Visual working memory influences the performance in virtual image-guided surgical intervention. *Surg Endosc* 2007; **21**: 2044-50
180 Luursema J, Buzink S, Verwey W, Jakimowicz J. Visuo-spatial ability in colonoscopy simulator training. *Adv Health Sci Educ* 2010; **15**: 685-94
181 Brandt M, Davies E. Visual-spatial ability, learning modality and surgical knot tying. *Can J Surg* 2006; **49**: 412-6
182 Gallagher A, Cowie R, Crothers I, Jordan-Black J, Satava R. PicSOr: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surg Endosc* 2003; **17**: 1468-71
183 Hedman L, Strom P, Andersson P, Kjellin A, Wredmark T, Fellander-Tsai L. High-level visual-spatial ability for novices correlates with performance in a visual-spatial complex surgical simulator task. *Surg Endosc* 2006; **20**: 1275-80
184 Wanzel K, Hamstra S, Anastakis D, Matsumoto E, Cusimano M. Effect of visual-spatial ability on learning of spatially-complex surgical skills. *Lancet* 2002; **359**: 230-1

185 Westman B, Ritter E, Kjellin A, et al. Visuospatial abilities correlate with performance of senior endoscopy specialist in simulated colonoscopy. *J Gastrointest Surg* 2006; **10**: 593-9
186 Van Herzeele I, O'Donoghue K, Aggarwal R, Vermassen F, Darzi A, Cheshire N. Visuospatial and psychomotor aptitude predicts endovascular performance of inexperienced individuals on a virtual reality simulator. *J Vasc Surg* 2010; **51**: 1035-42
187 Sweeney K, Hayes J, Chiavaroli N. Does spatial ability help the learning of anatomy in a biomedical science course? *Anat Sci Ed* 2014; **7**: 289-94
188 Lufler R, Zumwalt A, Romney C, Hoagland T. Effect of visual-spatial ability on medical students' performance in a gross anatomy course. *Anat Sci Ed* 2012; **5**: 3-9
189 Hassan I, Gerdes B, Koller M, et al. Spatial perception predicts laparoscopic skills on virtual reality laparoscopy simulator. *Child Nerv Syst* 2007; **23**: 685-9
190 Gettman M, Kondraske G, Traxer O, et al. Assessment of basic human performance resources predicts operative performance of laparoscopic surgery. *J Am Coll Surgeons* 2003; **197**: 489-96
191 Dashfield A, Smith J. Correlating fibreoptic nasotracheal endoscopy performance and psychomotor aptitude. *Br J Anaesth* 1998; **81**: 687-91
192 Keehner M, Lippa Y, Montello D, Tendick F, Hegarty M. Learning a spatial skill for surgery: How the contributions of abilities change with practice. *Appl Cognitive Psych* 2006; **20**: 487-503
193 Keehner M, Tendick F, Meng M, et al. Spatial ability, experience, and skill in laparoscopic surgery. *Am J Surg* 2004; **188**: 71-5
194 Wanzel K, Hamstra S, Caminiti M, Anastakis D, Grober E, Reznick R. Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. *Surgery* 2003; **134**: 750-7
195 Smith H, Kopp S, Johnson R, Long T, Cerhan J, Hebl J. Looking into Learning: Visuospatial and Psychomotor Predictors of Ultrasound-Guided Procedural Performance. *Reg Anesth Pain Med* 2012; **37**: 441-7
196 Shafqat A, Ferguson E, Thanawala V, Bedforth N, Hardman J, McCahon R. Visuospatial Ability as a Predictor of Novice Performance in Ultrasound-guided Regional Anesthesia. *Anesthesiology* 2015; **123**: 1188-97
197 Luursema J, Verwey W, Burie R. Visuospatial ability factors and performance variables in laparoscopic simulator training. *Learn Individ Differ* 2012; **22**: 632-8
198 Peters M, Laeng B, Latham K, Jackson M, Zaiyouna R, Richardson C. A redrawn Vandenberg and Kuse mental rotations test: different versions and factors that affect performance. *Brain Cogn* 1995; **28**: 39-58
199 Peters M. Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain Cogn* 2005; **57**: 176-84

200 O'Connor H, McGraw R. Clinical skills training: developing objective assessment instruments. *Med Educ* 1997; **31**: 359-63
201 Morgan P, Cleave-Hogg D, Guest C. A Comparison of Global Ratings and Checklist Scores from an Undergraduate Assessment Using an Anesthesia Simulator. *Acad Med* 2001; **76**: 1053-5
202 Ma I, Zalunardo N, Pachev G, et al. Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. *Adv Health Sci Educ* 2012; **17**: 457-70
203 Murray D, Boulet J, Avidan M, et al. Performance of Residents and Anesthesiologists in a Simulation-based Skill Assessment. *Anesthesiology* 2007; **107**: 705-13
204 Murray D, Boulet J, Kras J, McAllister J, Cox T. A Simulation-Based Acute Skills Performance Assessment for Anesthesia Training. *Anesth Analg* 2005; **101**: 1127-34
205 Murray D, Boulet J, Kras J, Woodhouse J, Cox T, McAllister J. Acute care skills in anesthesia practice: a simulation-based resident performance assessment. *Anesthesiology* 2004; **101**: 1084-95
206 Cunnington J, Neville A, Norman G. The risks of thoroughness: Reliability and validity of global ratings and checklists in an OSCE. *Adv Health Sci Educ* 1996; **1**: 227-33
207 Regehr G, MacRae H, Reznick R, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998; **73**: 993-7
208 Ilgen J, Ma I, Hatala R, Cook D. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ* 2015; **49**: 161-73
209 Martin J, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997; **84**: 273-8
210 Sivarajan M, Lane P, Miller E, et al. Performance Evaluation: Continuous Lumbar Epidural Anesthesia Skill Test. *Anesth Analg* 1981; **60**: 543-7
211 Friedman Z, Katznelson R, Devito I, Siddiqui M, Chan V. Objective assessment of manual skills and proficiency in performing epidural anesthesia--video-assisted validation. *Reg Anesth Pain Med* 2006; **31**: 304-10
212 Naik V, Matsumoto E, Houston P, et al. Fiberoptic orotracheal intubation on anesthetized patients: do manipulation skills learned on a simple model transfer into the operating room? *Anesthesiology* 2001; **95**: 343-8
213 Sultan S, Iohom G, Saunders J, Shorten G. A clinical assessment tool for ultrasound-guided axillary brachial plexus block. *Acta Anaesth Scand* 2012; **56**: 616-23
214 Cheung J, Chen E, Darani R, McCartney C, Dubrowski A, Awad I. The creation of an objective assessment tool for ultrasound-guided regional anesthesia using the Delphi method. *Reg Anesth Pain Med* 2012; **37**: 329-33

215 Wong D, Watson M, Kluger R, et al. Evaluation of a task-specific checklist and global rating scale for ultrasound-guided regional anesthesia. *Reg Anesth Pain Med* 2014; **39**: 399-408
216 Burckett-St-Laurent D, Niazi A, Cunningham M, et al. A Valid and Reliable Assessment Tool for Remote Simulation-Based Ultrasound-Guided Regional Anesthesia. *Reg Anesth Pain Med* 2014; **39**: 496-501
217 Anastakis D, Regehr G, Reznick R, et al. Assessment of technical skills transfer from the bench training model to the human model. *Am J Surg* 1999; **177**: 167-70
218 Harden R, Stevenson M, Downie W, Wilson G. Assessment of clinical competence using objective structured examination. *BMJ* 1975; **1**: 447-51
219 Harden R, Gleeson F. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979; **13**: 41-54
220 Hastie M, Spellman J, Pagano P, Hastie J, Egan B. Designing and Implementing the Objective Structured Clinical Examination in Anesthesiology. *Anesthesiology* 2014; **120**: 196-203
221 Bromley L. The Objective Structured Clinical Exam - practical aspects. *Curr Opin Anaesth* 2000; **13**: 675-8
222 Corrie K, Wiles M, Flack J, Lamb J. A novel objective structured clinical examination for the assessment of transfusion practice in anaesthesia. C*lin Teach* 2011; **8**: 97-100
223 Harden R. Revisiting 'Assessment of clinical competence using an objective structured clinical examination (OSCE)'. *Med Educ* 2016; **50**: 376-9
224 Goff B, Mandel L, Lentz G, et al. Assessment of resident surgical skills: is testing feasible? *Am J Obstet Gynecol* 2005; **192**: 1331-8; discussion 8-40
225 van der Vleuten C, Swanson D. Assessment of clinical skills with standardized patients: State of the art. *Teach Learn Med* 1990; **2**: 58-76
226 Swanson D, van der Vleuten C. Assessment of Clinical Skills With Standardized Patients: State of the Art Revisited. *Teach Learn Med* 2013; **25**: S17-S25
227 Ben-Menachem E, Ezri T, Ziv A, Sidi A, Brill S, Berkenstadt H. Objective Structured Clinical Examination-based assessment of regional anesthesia skills: the Israeli National Board Examination in Anesthesiology experience. *Anesth Analg* 2011; **112**: 242-5
228 Berkenstadt H, Ziv A, Gafni N, Sidi A. The validation process of incorporating simulation-based accreditation into the anesthesiology Israeli national board exams. *Isr Med Assoc J* 2006; **8**: 728-33
229 Berkenstadt H, Ziv A, Gafni N, Sidi A. Incorporating simulation-based objective structured clinical examination into the Israeli National Board Examination in Anesthesiology. *Anesth Analg* 2006; **102**: 853-8

230 Naeem N. Validity, reliability, feasibility, acceptability and educational impact of direct observation of procedural skills (DOPS). *J Coll Physicians Surg Pak* 2013; **23**: 77-82
231 Mitchell C, Bhat S, Herbert A, Baker P. Workplace-based assessments of junior doctors: do scores predict training difficulties? *Med Educ* 2011; **45**: 1190-8
232 The Royal College of Anaesthetists. Changes to Assessments for Specialty Trainees. 2015. Available from http://www.rcoa.ac.uk/news-and-bulletin/rcoa-news-and-statements/changes-assessments-specialty-trainees (accessed 10 March 2016
233 Woolliscroft J, Stross J, Silva J. Clinical competence certification: a critical appraisal. *J Med Ed* 1984; **59**: 799-805
234 Kroboth F, Hanusa B, Parker S, et al. The inter-rater reliability and internal consistency of a clinical evaluation exercise. *J Gen Intern Med* 1992; **7**: 174-9
235 Norcini J, Blank L, Arnold G, Kimball H. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med* 1995; **123**: 795-9
236 Hawkins R, Margolis M, Durning S, Norcini J. Constructing a validity argument for the mini-Clinical Evaluation Exercise: a review of the research. *Acad Med* 2010; **85**: 1453-61
237 Norcini J, Blank L, Duffy F, Fortna G. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med* 2003; **138**: 476-81
238 Alves de Lima A, Barrero C, Baratta S, et al. Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. *Med Teach* 2007; **29**: 785-90
239 Golnik K, Goldenhar L. The ophthalmic clinical evaluation exercise: reliability determination. *Ophthalmology* 2005; **112**: 1649-54
240 Durning S, Cation L, Markert R, Pangaro L. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Acad Med* 2002; **77**: 900-4
241 Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *Br J Anaesth* 2003; **90**: 580-8
242 Yule S, Paterson-Brown S. Surgeons' non-technical skills. *Surg Clin N Am* 2012; **92**: 37-50
243 Flin R, Patey R. Non-technical skills for anaesthetists: developing and applying ANTS. *Best Pract Res Clin Anaesth* 2011; **25**: 215-27
244 Jepsen R, Dieckmann P, Spanager L, et al. Evaluating structured assessment of anaesthesiologists' non-technical skills. *Acta Anaesth Scand* 2016; **60**: 756-66.
245 Watson M, Wong D, Kluger R, et al. Psychometric evaluation of a direct observation of procedural skills assessment tool for ultrasound-guided regional anaesthesia. *Anaesthesia* 2014; **69**: 604-12

246 Chuan A, Thillainathan S, Graham P, et al. Reliability of the Direct Observation of Procedural Skills assessment tool for ultrasound-guided regional anaesthesia. *Anaesth Intens Care* 2016; **44**: 201-8

247 Roberts L. Assessing anaesthesia trainees at work: opportunities and challenges. *Anaesth Intens Care* 2016; **44**: 194-7

248 Garden A. Reliability of numerical scales used for direct observation of procedural skills. *Anaesth Intens Care* 2016; **44**: 198-200

249 Chuan A, Graham P, Wong D, et al. Design and validation of the Regional Anaesthesia Procedural Skills Assessment Tool. *Anaesthesia* 2015; **70**: 1401-11

250 Chuan A, Graham P, Forrest K, et al. Regional anaesthesia assessment tools – a reply. *Anaesthesia* 2016; **71**: 473-4

251 Chuan A, Lim Y, Aneja H, et al. A randomised controlled trial comparing meat-based with human cadaveric models for teaching ultrasound-guided regional anaesthesia. *Anaesthesia* 2016; **71**: 921-9.

252 Gupta AK, Morton JR. Cadaveric training – the solution for ultrasound-guided regional anaesthesia? *Anaesthesia* 2016; **71**: 874-8.

253 Duce N, Gillett L, Descallar J, Tran M, Siu S, Chuan A. Visuospatial ability and novice brachial plexus sonography performance. *Acta Anaesth Scand* 2016; **60**: 1161-9.