# The utility of high-stakes technical performance assessments in surgical training: establishing a framework across the continuum from selection to certification

*Conor Toale MB BCh BAO(Hons.), MSc, MRCSI*
*Department of Surgical Affairs, Royal College of Surgeons in Ireland*

~

Submitted to the School of Postgraduate Studies, University of Medicine and Health Sciences at the Royal College of Surgeons in Ireland, in fulfillment of the degree of Doctor of Philosophy

Supervised by

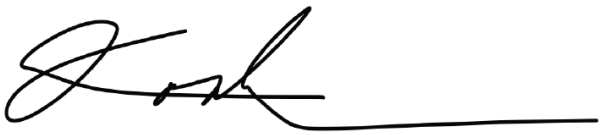
Mr. Dara Kavanagh MCh FRCSI
Head of Surgical Research, Department of Surgical Affairs, RCSI

and

Dr. Marie Morris PGDipHEd MSc PhD
Senior Lecturer, Department of Surgical Affairs, RCSI

**Candidate Thesis Declaration**

I declare that this thesis, which I submit to RCSI for examination in consideration of the award of a doctorate of philosophy (PhD) is my own personal effort. Where any of the content presented is the result of input or data from a related collaborative research programme this is duly acknowledged in the text such that it is possible to ascertain how much of the work is my own. I have not already obtained a degree in RCSI or elsewhere on the basis of this work. Furthermore, I took reasonable care to ensure that the work is original, and, to the best of my knowledge, does not breach copyright law, and has not been taken from other sources except where such work has been cited and acknowledged within the text.

Signed: 

Student Number:

20107481

Date:

26/09/23

## Intellectual Property Declaration

The contents of the enclosed manuscript are confidential and should not be disclosed, or disseminated in any way, to any third party other than to staff or students of the Royal College of Surgeons in Ireland or an external examiner appointed for the purpose of reviewing the manuscript.

## Table of Contents


| | Page |
|---|---|
| Thesis declaration | 2 |
| Intellectual property declaration | 3 |
| List of abbreviations | 10 |
| List of tables | 13 |
| List of figures | 14 |
| Acknowledgments | 16 |
| Dedication | 18 |
| Summary | 19 |
| Publications and presentations | 21 |
| List of published scientific papers for consideration in this thesis | 25 |
| Related published scientific papers | 28 |
| Related unpublished supporting scientific papers | 30 |
| **Chapter 1: Introduction: The current status of high-stakes technical performance assessments in surgical training** | 31 |
| Abstract | 32 |
| 1.1 Introduction | 34 |
| 1.2 Methods | 35 |
| 1.3 Results | 36 |
| 1.3.1. Resident selection or recruitment | 41 |
| 1.3.1.1 Psychomotor assessments | 41 |
| 1.3.1.2 Simulated operative tasks | 41 |
| 1.3.1.3 Implementation and Feasibility | 42 |
| 1.3.1.4 Acceptability | 42 |
| 1.3.2 Progression through training and remediation | 43 |
| 1.3.2.1 Progression to the next training year | 43 |
| 1.3.2.2 Progression from Core Training to Higher Specialist Training | 43 |
| 1.3.2.3 Progression to fellowship training | 44 |
| 1.3.2.4 Acceptability of SBA used to inform progression through training | 45 |
| 1.3.3 Entrustment, credentialing, and procedure-specific certification | 45 |
| 1.3.3.1 Validity and Implementation | 45 |
| 1.3.3.2 Acceptability | 46 |
| 1.3.4 Certification | 46 |
| 1.3.4.1 Mandatory in-training assessments | 46 |
| 1.3.4.2 End of training certification | 47 |
| 1.3.4.3. Acceptability | 48 |
| 1.4 Discussion | 49 |
| 1.5 Conclusion | 52 |
| 1.6 Thesis objectives and research questions | 53 |
| 1.7 Data sources for this thesis | 60 |
| **Chapter 2: Determining the need for, and acceptability of, high-stakes technical performance assessments in surgical training** | 62 |
| 2.1 Training to Proficiency in Surgery using Simulation: Is there a moral obligation? | 65 |

2.2. Assessing Operative Skill in the Competency-Based Education Era: Lessons from the UK and Ireland 66
2.3 Defining Operative Experience Targets in Surgical Training: A Systematic Review 67
2.4 Perceptions and Experiences of Simulation-Based Assessment of Technical Skill in Surgery: A Scoping Review 68
2.5 Perspectives on simulation-based assessment of operative skill in surgical training 69
*2.6. Additional Material: The Acceptability of Assessing Operative Competence Outside of the Operating Room: A Survey of Surgical Trainees* *70*
*Abstract* *71*
*2.6.1 Introduction* *73*
*2.6.2 Methods* *74*
*2.6.3 Results* *75*
*2.6.4 Discussion* *83*
**Chapter 3: Validity evidence for the use of objective technical aptitude assessments at the time of surgical trainee selection** 86
3.1 Baseline assessments of psychomotor aptitude are associated with the future operative performance of surgical trainees 87
*3.2 Additional Material: Performance at National Selection and Future Performance in Surgical Training* 88
**Chapter 4: Exploring the utility of an implemented simulation-based assessment designed to inform decisions on progression from Core Surgical Training to Higher Specialist training** 89
4.1 Factors negatively impacting upon the reliability of workplace-based assessments of operative competence 91
4.2 Exploring the Number of Workplace Observations Required to Reliably Assess Operative Competence in Core Surgical Training 92
*4.3 Additional Material: Prospective Validation of the Operative Surgical Skill (OSS) Examination: A Simulation-Based Assessment of Operative Competence for Core Surgical Trainees* 93
*Abstract* 94
*4.3.1 Introduction* 95
*4.3.2 Methods* 96
*4.3.2.1 Study design, setting and participants* 96
*4.3.2.2 Study outcome measures* 97
*4.3.2.3 Statistical analysis* 99
*4.3.3 Results* 99
*4.3.3.1 Internal structure validity* 100
*4.3.3.2 Consequential validity* 102
*4.3.4 Discussion* 103
*4.3.5 Conclusion* 105
*4.4 Additional Material: Comparing the Reliability of Simulation-Based and Workplace-Based Assessments of Operative Competence in Core Surgical Training* 106
*Abstract* 107
*4.4.1 Introduction* 108

*4.4.2 Methods* 109
*4.4.2.1 Study design, setting and participants* 109
*4.4.2.2 Data sources/ measurement* 109
*4.4.2.3 Statistical methods* 111
*4.4.3 Results* 112
*4.4.3.1 Relative contributions or trainee ability and procedure performed to score variance* 112
*4.4.3.2 Decision studies* 113
*4.4.4 Discussion* 115
*4.4.5 Conclusion* 117
**Chapter 5: Development of a novel simulation-based assessment designed to evaluate operative competence in senior general surgery trainees** 118
5.1 Generating a prioritised list of operative procedures for simulation-based assessment of General Surgery Trainees through consensus 120
5.2 Training and Assessment using the LapSim Laparoscopic Simulator: A Scoping Review of Validity Evidence 121
*5.3. Additional Material: Development and Validation of a Simulation-Based Assessment of Operative Competence for Higher Specialist Trainees in General Surgery* 122
*Abstract* 123
*5.3.1 Introduction* 125
*5.3.2 Methods* 126
*5.3.2.1 Initial development of the assessment framework* 126
*5.3.2.2 Assessment Instruments* 127
*5.3.2.3 Participants* 128
*5.3.2.4 Procedure* 128
*5.3.2.5 Statistical Analysis* 129
*5.3.3. Results* 130
*5.3.3.1 Content Validity* 131
*5.3.3.2 Internal structure validity* 131
*5.3.3.3 Relationships with other variables* 132
*5.3.3.4 Response process validity* 135
*5.3.3.5 Consequences* 135
*5.3.3.6 Feasibility* 136
*5.3.4 Discussion* 136
*5.2.5 Conclusion* 139
*5.4. Additional Material: Using simulation to assess operative competence: a qualitative case study of a novel simulation-based assessment* 140
*Abstract* 141
*5.4.1 Introduction* 143
*5.4.2 Methods* 144
*5.4.2.1 Context* 144
*5.4.2.2 Study design* 144
*5.4.2.3 Reflexivity* 146

*5.4.2.4 Analysis and findings* 146
*5.4.3 Results* 147
*5.4.3.1 Simulation as an acceptable modality to assess canonical competence* 148
*5.4.3.2 The authenticity gap – higher levels of operative competence are better assessed in the workplace* 149
*5.4.3.3 Concerns regarding the implications of sim-based operative competence assessment* 151
*5.4.3.4 Bridging the gap: developing simulation-based assessment to assess operative competence* 152
*5.4.3.5 Acknowledging the gap: embracing the complementarity of workplace-based assessment and simulation-based assessment* 153
*5.4.4 Discussion* 154
*5.4.5 Conclusion* 156
*5.5 Additional material: Why do residents fail simulation-based assessments of operative competence? A qualitative analysis* 157
**Chapter 6: General Discussion** 158
6.1 Main Findings 159
6.1.1 Determining the need for simulation-based technical skill assessment in surgical training, in the context of a multi-modality programmatic approach to procedural assessment 159
6.1.2 Validity evidence for the use of objective technical aptitude assessments at the time of surgical trainee selection 160
6.1.3 Exploring the utility of an implemented simulation-based assessment designed to inform decisions on progression from Core Surgical Training to Higher Specialist training 161
6.1.4 Development of a novel simulation-based assessment designed to evaluate operative competence in senior general surgery trainees 162
6.2 Incorporating Simulation-Based Assessment of Operative Competence into the Certification of Trainees: A Proposed Blueprint 163
Abstract 164
6.2.1 Competence, Competency-Based Education, and Surgery 166
6.2.2 Operative assessment in surgery 167
6.2.3 Simulation, assessment and certification 168
6.2.4 Professional authenticity and simulation-based assessment of operative competence 169
6.2.5 The simulated and workplace environments; two views on ‘construct under-representation’ 172
6.2.6 High acuity, low occurrence procedures in surgery 173
6.2.7 Simulation-based procedural assessment and its role in certification 174
6.2.8 Conclusion 175
6.3 High-stakes technical performance assessments across the continuum of surgical training 176
Abstract 177

6.3.1 Surgical Skill and the Acquisition of Operative Competence 178
6.3.2 Mastery Learning, Competence and Surgical Training 179
6.3.3 Assessment across the Continuum of Surgical Education 181
6.3.4 Assessing technical skill outside of the operating room: representing the appropriate construct 183
6.3.5 Technical Performance assessments across the continuum of surgical training 184
6.3.6 Proposals regarding the use of technical performance assessments in surgical training 186
6.3.6.1 Selection and Recruitment 186
6.3.6.2 Progression through training 187
6.3.6.3 Certification 188
6.3.7 Looking to the future 188
6.3.8 Conclusions 189
6.4 Future Directions 191
6.4.1 Using technical aptitude assessments to inform selection decisions 191
6.4.2. Using SBA to inform progression through training 191
6.4.3. Using SBA to inform end-of-training certification decisions 192
6.5 Methodological Considerations 193
6.5.1 Strengths 193
6.5.2 Limitations 195
6.6 Conclusion 198
7. Bibliography 200
*8. Supplemental Material* *245*
*8.1 Supplemental Material – Chapter 1* *245*
*8.1.1 Appendix 1.1* *245*
*8.1.2 Appendix 1.2* *250*
*8.2 Supplemental Material – Chapter 2* *251*
*8.2.1 Appendix 2.2.1* *251*
*8.2.2 Appendix 2.2.2* *252*
*8.2.3 Appendix 2.3.1* *253*
*8.2.4 Appendix 2.3.2* *254*
*8.2.5 Appendix 2.3.3* *255*
*8.2.6 Appendix 2.4.1* *256*
*8.2.7 Appendix 2.4.2* *257*
*8.2.8 Appendix 2.5.1* *258*
*8.2.9 Appendix 2.5.2* *259*
*8.2.10 Appendix 2.6.1* *260*
*8.3 Supplemental Material – Chapter 3* *266*
*8.3.1 Appendix 3.1.1* *266*
*8.3.2 Appendix 3.2.1* *269*
*8.4 Supplemental Material – Chapter 4* *275*
*8.4.1 Appendix 4.1.1* *275*
*8.4.2 Appendix 4.2.1* *278*
*8.4.3 Appendix 4.2.2* *281*

*8.4.4 Appendix 4.2.3* *282*
*8.4.5 Appendix 4.2.4* *283*
*8.4.6 Appendix 4.3.1* *284*
*8.4.7 Appendix 4.4.1* *285*
*8.5 Supplemental Material – Chapter 5* *288*
*8.5.1 Appendix 5.1.1* *288*
*8.5.2 Appendix 5.1.2* *289*
*8.5.3 Appendix 5.2.1* *290*
*8.5.4 Appendix 5.2.2* *291*
*8.5.5 Appendix 5.3.1* *292*
*8.5.6 Appendix 5.3.2* *294*
*8.5.6 Appendix 5.3.3* *295*
*8.5.7 Appendix 5.3.4* *296*
*8.5.8 Appendix 5.3.5* *297*
*8.5.9 Appendix 5.3.6* *299*
*8.5.10 Appendix 5.3.7* *301*
*8.5.11 Appendix 5.5.1* *303*
9. Corresponding Author Declarations 306

**List of Abbreviations:**

| | |
|---|---|
| AAA | Abdominal Aortic Aneurysm |
| ABCRS | American Board of Colon and Rectal Surgery |
| ABS | American Board of Surgery |
| ABSITE | American Board of Surgery In-Training Examination |
| ACGME | American College of Graduate Medical Education |
| ACLS | Advanced Cardiovascular Life Support |
| ACS | American College of Surgeons |
| AMEE | Association of Medical Education in Europe |
| ANZ | Australia and New Zealand |
| APDS | Association of Programme Directors in Surgery |
| ARCP | Annual Review of Competence Progression |
| ASMBS | American Society for Metabolic and Bariatric Surgery |
| ATIS | Adjusted Total Item Score |
| BLS | Basic Life Support |
| BMC | BioMed Central |
| BMJ | British Medical Journal |
| CA | Consultant Anaesthetist |
| CanMEDS | Canadian Medical Education Directives For Specialists |
| CAMEO | Clinical Assessment and Management Examination – Outpatient |
| CAMES | Copenhagen Academy for Medical Education and Simulation |
| CBD | Common Bile Duct |
| CBE | Competency-Based Education |
| CCC | Clinical Competency Committee |
| CCST | Certification of Completion of Specialist Training |
| CCT | Certification of Completion of Training |
| CERP | Conventional End-of Rotation Environment |
| CI | Confidence Interval |
| COREQ | Consolidated Criteria for Reporting Qualitative Research |
| COSATS | Colorectal Objective Structured Assessment of Technical Surgery |
| COVID | Corona Virus Disease |
| CS | Consultant Surgeon |
| CST | Core Surgical Trainee |
| CT | Core Trainee |
| DIEP | Deep Inferior Epigastric Perforator (flap) |
| DOI | Digital Object Identifier |
| DOPS | Direct Observation of Procedural Skills |
| DU | Duodenal Ulcer |
| EGS | Emergency General Surgery |
| ELAPE | Extra-Levator Abdomino-Perineal Excision |
| EM | Executive Management representative |
| EMBASE | Excerpta Medica Database |
| ENT | Ear, Nose and Throat (surgery) |
| EPA | Entrustable Professional Activity |
| ERCP | Endoscopic Retrograde Cholangio-Pancreatography |
| EUA | Examination under Anaesthesia |
| FLS | Fundamentals of Laparoscopic Surgery |

| | |
|---|---|
| FNA | Fine Needle Aspiration |
| FRCS | Fellowship of the Royal College of Surgeons |
| FRCSI | Fellowship of the Royal College of Surgeons in Ireland |
| GDPR | General Data Protection Regulation |
| GI | Gastro-Intestinal (Surgery) |
| GMC | General Medical Council |
| GOSATS | General Surgery Objective Structured Assessment of Technical Skill |
| GOSLE | Generic Operative Supervised Learning Event |
| GR | Global Rater |
| GSS | Global Summary Score |
| HIPE | Hospital In-Patient Enquiry System |
| HPB | Hepatobiliary (surgery) |
| HST | Higher Specialist Training |
| IBM | International Business Machines |
| ICS | Interpersonal/ Communication Skills |
| IMC | Irish Medical Council |
| IMV | Inferior Mesenteric Vein |
| ISCP | Intercollegiate Surgical Curriculum Programme |
| JAMA | Journal of the American Medical Association |
| JBI | Joanna Briggs Institute |
| MCQ | Multiple Choice Question |
| MD | Doctor of Medicine (degree) |
| MEDLINE | Medical Literature Analysis and Retrieval System Online |
| MINQUE | Minimum Norm Quadratic Unbiased Education |
| MK | Medical Knowledge |
| MRCSI | Member of the Royal College of Surgeons in Ireland |
| MS | Medical Student |
| MVI | Major Vessel Injury |
| NA | Not Applicable |
| NAF | Needs Assessment Formula |
| NHLBI | National Heart, Lung and Blood Institute (USA) |
| NI | Northern Ireland |
| NIH | National Institute for Health (USA) |
| NOTSS | Non-Technical Skills for Surgeons |
| NQAIS | National Quality Assurance and Improvement System |
| NR | Nursing Representative |
| NSQIP | National Surgical Quality Improvement Programme (USA) |
| NVIVO | NVivo is a qualitative data analysis computer software package produced by QSR International |
| OBM | Objective Borderline Method |
| OCAP | Orthopaedic Curriculum and Assessment Project |
| OHNS | Otolaryngology/ Head and Neck Surgery |
| OPRS | Operative Performance Rating System |
| OSATS | Objective Structured Assessment of Technical Skill |
| OSCE | Objective Structured Clinical Assessment |
| PBA | Procedure-Based Assessment |
| PBLI | Practice-Based Learning and Improvement |
| PC | Patient Care (milestone evaluation) |

| | |
|---|---|
| PGY | Post-Graduate Year |
| POP | Performance of Procedures (milestone evaluation) |
| PR | Patient Representative |
| PRISMA | Preferred Reporting Items for Systematic Reviews and Meta-Analyses |
| PROF | Professionalism (milestone evaluation) |
| QSR | QSR international (qualitative research software developer) |
| RACS | Royal Australian College of Surgeons |
| RCPSC | Royal College of Physicians and Surgeons of Canada |
| RCSI | Royal College of Surgeons in Ireland |
| RM | Risk Manager |
| RMO | Resident Medical Officer |
| SBA | Simulation-Based Assessment |
| SBP | Systems-Based Practice |
| SCORE | Surgical Council on Resident Education |
| SD | Standard Deviation |
| SEPA | Surgical Entrustable Professional Activity |
| SET | Surgical Education and Training |
| SHO | Senior House Officer |
| SIMPL | System for Improving and Measuring Procedural Learning |
| SKL | Skill' score |
| SLNB | Sentinel Lymph Node Biopsy |
| SPSS | Statistical Package for Social Sciences |
| SR | Simulation Representative |
| SSAOP | Supervised Structured Assessment of Operative Performance |
| ST | Surgical Trainee |
| STAR | Surgical Trainee Assessment and Readiness |
| STROBE | STrengthening the Reporting of Observational studies in Epidaemiology |
| TAR | Trainer Assessment Report |
| TRAM | Transverse Rectus Abdominus Myocutaneous |
| UK | United Kingdom |
| US | United States |
| USA | United States of America |
| VC | Score variance attributable to inter-case variation (reliability analysis) |
| VE | Score variance attributable to residual error (reliability analysis) |
| VR | Virtual reality |
| VS | Score variance attributable to the observing surgeon-assessor (reliability analysis) |
| VSC | Score variance interaction: variance attributable to individual surgeons observing individual cases (reliability analysis) |
| VT | Score variance attributable to an individual trainee (reliability analysis) |
| WBA | Workplace-Based Assessment |
| WHO | World Health Organisation |

## List of Tables

**Chapter 1**
*Table 1.1: Inclusion and Exclusion Criteria – The current status The current status of high-stakes technical performance assessments in surgical training* 36
*Table 1.2: Characteristics of included studies – the current status of simulation as a high-stakes technical skill assessment modality in surgical training* 38
**Chapter 2**
**Chapter 3**
**Chapter 4**
*Table 4.3.1: Scores for each station presented as mean ± standard deviation for each cohort (scored from 0 – 100).* 100
*Table 4.3.2: Internal consistency.* 100
*Table 4.3.3: Observed score variance attributable to inter-trainee variance, Var(trainee), inter-station variance, Var(Station), inter-cohort variance referring to CST vs MSSP candidates, Var(Cohort), the interaction between trainee and individual station variance, Var(trainee*station), and the residual variance attributable to error, Var(Error).* 101
*Table 4.4.1: Number of analysed assessments with mean checklist scores* 112
*Table 4.4.2: Contribution of trainee ability, procedure/ station variance, and trainee procedure-specific competence to observed score variance.* 113
**Chapter 5**
*Table 5.3.1: Mean scores awarded across assessment stations* 130
*Table 5.3.2: Cronbach's alpha* 131
*Table 5.3.3: Correlation (Pearson) between individual station scores and total scores awarded* 132
*Table 5.3.4: Mean score differences for 'Junior' vs. 'Senior' trainees across individual stations.* 134
*Table 5.3.5: Borderline regression passing scores and passing rates:* 136
**Chapter 6**

## List of Figures

**Chapter 1**

*Figure 1.1: PRISMA flow-chart, The current status of simulation as a high-stakes technical skill assessment modality in surgical training* 37

*Figure 1.6.1: Currently utilised and proposed technical performance assessments in Irish surgical training* 54

*Figure 1.6.2: Currently utilised assessment curriculum in General Surgery Higher Specialist Training.* 55

*Figure 1.6.3: Thesis conceptualised* 57

**Chapter 2**

*Figure 2.6.1: Limitations of SBA, and methods for promoting its acceptable use in surgical training, generated from thematic analysis of open-response survey items* 83

**Chapter 3**

**Chapter 4**

*Figure 4.3.1: Decision study. Increasing G-coefficient with increasing numbers of hypothetical stations.* 101

*Figure 4.3.2: Borderline Regression analyses for each station.* 103

*Figure 4.4.1: Reliability co-efficient (G) of OSCEs and in-theatre (SSAOP) assessments with increasing numbers of stations (OSCE) or procedures observed (SSAOP).* 114

**Chapter 5**

*Figure 5.3.1: Difference in mean scores achieved by 'Junior' (ST2 – ST4) trainees, and 'Senior' (ST5 – ST8) trainees.* 133

*Figure 5.3.2: Relationship between training year and mean score obtained* 133

*Figure 5.3.3: Generation of a pass-fail benchmark* 135

*Figure 5.4.1: Generated themes* 147

**Chapter 6**

*Figure 6.2.1: Assessments of increasing professional authenticity (A) capture a broader range of knowledge, skills and attitudes required to represent the construct of 'competence' (B), in turn leading to increasing validity when assessing competence along the continuum from canonical competence through contextual competence to personalized competence (C).* 171

*Figure 6.3.1: A simplified representation of the Irish surgical training curricula.* 182

*Figure 6.3.2: The professional authenticity of assessments should increase along the continuum of surgical training from selection to end-of-training certification* 185

*Figure 6.3.3: Current simulation technology limits the ability to assess contextual competence due to deficiencies in fidelity and therefore professional authenticity* 185

*Figure 6.3.4: Framework for the use of technical performance assessments across surgical training, with reference to a proposed assessment framework for robotic surgical training across the continuum from selection to certification* 189

**Acknowledgements:**

I would like to thank the following people for their support in carrying out this research:

I am deeply indebted to my inspirational supervisors Mr. Dara Kavanagh and Dr. Marie Morris for their unwavering enthusiasm, support and guidance throughout this project. It has been an absolute pleasure exploring the topics presented in this thesis with you. You have allowed me to pursue my own line of independent enquiry and supported every research aspiration of mine, assisting with every big and small step. All the while, you managed to make this experience rewarding and enriching for me – both academically and beyond. Lastly, you made the journey fun; I hope we can continue our research relationship long into the future.

I would also express my deepest appreciation to the academic, technical and support staff within the department of Surgical Affairs, the department of Simulation, and the Quality Enhancement Office at the Royal College of Surgeons in Ireland. In particular, Prof. Oscar Traynor, Ms. Dara O'Keeffe, Prof. Eva Doherty, Ms. Leonie Heskin, Mr. Adam Roche, Mr. Miroslav Voborsky, Ms. Sara Gross, Ms. Jane Cunningham, Ms. Sinead Dixon, Ms. Paula Mansell, Ms. Adina Orac and Ms. Joanna Zawadska deserve special thanks for their invaluable support with this work. The wisdom, intellectual rigour and experience of Dr. Fiona Boland from the Data Science Centre has contributed hugely to the statistical analyses presented herein; I have learnt a great deal from working with her. Prof. Judith Strawbridge from the School of Pharmacy in RCSI has helped me to navigate the previously unfamiliar waters of qualitative research, for which I am extremely grateful.

I thank the Hermitage Medical Clinic for funding this research as part of the Strategic Academic Recruitment (StAR) programme, and for supporting me professionally during the first two years of my academic research time. I would like to further thank the StAR MD team and School of Postgraduate Studies for guiding me through the

PhD process. In particular, Professor Emer Reeves and Kathleen Molloy were incredibly generous with their time and expertise.

I would like extend my sincere thanks to the surgeons, surgical trainees and other medical and non-medical representatives for their engagement as participants or assessors throughout this project. Asking for their time, expertise and patience is no small request given the pressures they face clinically and academically every day. Their enthusiasm for the research interest of a junior colleague will not be forgotten.

Finally, I would like to thank my family and friends. To my fiancée Aideen Fallon, thank you for listening to, reading, and critiquing my work at all hours of the day and night. To my mum, Siobhan, and dad, Pearse, thank you for your love and support throughout my life – producing thesis would not have been possible without your help.

## Dedication

This thesis is dedicated to my grandmothers Tess Hennessy and Maura Toale, and my grandfather Jim Hennessy, who passed away before they could see its completion.

## Summary

The traditional Halstedian of apprenticeship training in surgery has undergone significant revolution in recent decades [1]. Evolving patient expectations regarding the role of surgical trainees in their care[2], an increased emphasis on theatre efficiency [3], and well explored concerns regarding perceived operative competence and confidence of graduating trainees [4-6], has led to a re-evaluation of the training paradigm. 'Competency-based' approaches to outcome-driven training and assessment are well established across training jurisdictions[7 8], leading to the development of nominally time-independent postgraduate surgical programs [9]. The surgical specialties pose a unique challenge in this regard, due to the requirement for robust and reliable methods of teaching and assessing competence in operative skill [10]. Determinations of competence, in the context of the surgical training programme of Ireland and the United Kingdom, are made at Annual Review of Competence Progression (ARCP) meetings and are informed by performance in workplace-based assessments (WBAs), operative numbers (as recorded by trainees in an electronic logbook), and educational supervisor reports. From August of 2021, surgical training in this jurisdiction has moved to an explicitly outcome-based curriculum[11]. Trainees are measured against high level outcomes - 'Capabilities in Practice' and 'Generic Professional Capabilities'. This training programme is nominally time-independent[11]. In order to transition to a truly competency-based paradigm, operative skill will need to be assessed and certified through an objective, standardised, and valid approach.

The role of simulation in surgery has evolved in tandem with competency-based education[12]. Simulation is commonplace in surgical training curricula [13], and is increasingly being deployed as a method of assessing a trainee's competence [14 15]. Studies have demonstrated the safe transfer of simulator-acquired operative skill to the operating theatre [16]. Whether simulation-training ultimately translates to improved patient safety and surgical outcome measures has not yet been firmly established [17]. Simulation has most notably been used in high-stake assessments

through the implementation of the Colorectal Objective Structured Assessment of Technical Skill by the American Board of Colon and Rectal Surgery[14 18]. Such assessments have not been used for higher specialist training in the UK and Ireland to date. Beyond end-of-training certification, simulation (or other technical performance assessments) could be used as a measure of skill or competence at a number of key checkpoints or high-stakes scenarios in surgery, namely: trainee selection, progression through training, and autonomy granting.

This thesis will seek to explore the use of high-stakes technical performance assessments of operative skill and competence in surgical training. The validity, acceptability and reliability of technical performance assessments will be explored, using both quantitative and qualitative methods, at three key time points: trainee selection, in-training progression, and competence certification in senior training. For the purpose of this thesis, 'performance assessments' are defined as assessments that take place outside of the workplace context, in comparison to 'workplace-based assessments'. The validity of psychomotor assessments of fundamental technical aptitude will be examined at entry to Core Surgical Training. The validity of simulation-based assessment will be assessed during Core Surgical Training. Finally, we will seek to develop and validate a simulation-based assessment curriculum suitable for the assessment of operative competence in senior Higher Specialist trainees in General Surgery.

Findings from this work will be used to inform a blueprint, or proposed framework, regarding the use of technical performance assessments in high stakes scenarios surgical training, from trainee selection to certification as an independent practitioner.

## Publications and Presentations

*Publications:*

Toale C, Morris M, Kavanagh D. Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review. Am J Surg. 2021 Oct;222(4):723-730. doi: 10.1016/j.amjsurg.2021.03.035. Epub 2021 Apr 6. PMID: 33849711.

Toale C, Nally DM, Ryan DM, Morris M, Kavanagh DO. How do Trainers and Trainees use Formative Workplace-based Assessments of Operative Competence to Deliver and Receive Structured Feedback? J Surg Educ. 2022 Mar-Apr;79(2):485-491. doi: 10.1016/j.jsurg.2021.08.023. Epub 2021 Sep 28. PMID: 34593328.

Toale C, Morris M, Kavanagh DO. Training to proficiency in surgery using simulation: is there a moral obligation? J Med Ethics. 2022 Jan 6:medethics-2021-107678. doi: 10.1136/medethics-2021-107678. Epub ahead of print. PMID: 34992083.

Toale C, Morris M, Kavanagh DO. Assessing Operative Skill in the Competency-based Education Era: Lessons From the UK and Ireland. Ann Surg. 2022 Apr 1;275(4):e615-e625. doi: 10.1097/SLA.0000000000005242. PMID: 35129531.

Toale C, Morris M, Ryan DM, Boland F, Doherty EM, Traynor OJ, Kavanagh DO. Baseline Assessments of Psychomotor Aptitude are Associated with the Future Operative Performance of Surgical Trainees A Prospective Observational Cohort Study. Ann Surg. 2022 Jul 15. doi: 10.1097/SLA.0000000000005565. Epub ahead of print. PMID: 35837886.

Toale C, O'Byrne A, Morris M, Kavanagh DO. Defining operative experience targets in surgical training: A systematic review. Surgery. 2022 Nov;172(5):1364-1372. doi: 10.1016/j.surg.2022.07.015. Epub 2022 Aug 26. PMID: 36038374.

Toale C, Morris M, Kavanagh DO. Training and assessment using the LapSim laparoscopic simulator: a scoping review of validity evidence. Surg Endosc. 2022 Sep 19. doi: 10.1007/s00464-022-09593-0. Epub ahead of print. PMID: 36123545.

Toale C, Morris M, Kavanagh DO. Perspectives on simulation-based assessment of operative skill in surgical training. Med Teach. 2022 Oct 26:1-8. doi: 10.1080/0142159X.2022.2134001. Epub ahead of print. PMID: 36288727.

Toale C, Morris M, Kavanagh DO. Comment on: The impact of virtual-reality simulation training on operative performance in laparoscopic cholecystectomy: meta-analysis of randomized clinical trials. BJS Open. 2022 Nov 2;6(6):zrac134. doi: 10.1093/bjsopen/zrac134. PMID: 36412113.

Toale C, O'Byrne A, Morris M, Kavanagh DO. Characterizing individual trainee learning curves in surgical training: Challenges and opportunities. Surgeon. 2022. doi: 10.1016/j.surge.2022.11.003

Toale C, Morris M, Kavanagh DO. Perspectives on simulation-based assessment of operative skill in surgical training. Med Teach. 2022 Oct 26:1-8. doi: 10.1080/0142159X.2022.2134001. Epub ahead of print. PMID: 36288727.

Toale C, Morris M, O'Keeffe D, Boland F, Ryan DM, Nally DM, Kavanagh DO. Assessing operative competence in core surgical training: A reliability analysis. Am J Surg. 2023 Jun 30:S0002-9610(23)00273-8. doi: 10.1016/j.amjsurg.2023.06.020. Epub ahead of print. PMID: 37481408.

Toale, C., Morris, M. & Kavanagh, D.O. Why do residents fail simulation-based assessments of operative competence? A qualitative analysis. Global Surg Educ 2, 83 (2023). https://doi.org/10.1007/s44186-023-00161-1

Toale C, Morris M, Konge L, Nayahangan LJ, Roche A, Heskin L, Kavanagh DO. Generating a prioritised list of operative procedures for simulation-based assessment

of General Surgery Trainees through consensus (Accepted for publication, Annals of Surgery)

*Presentations:*

Development and Validation of a Novel Simulation-Based Assessment of Competence for Higher Specialist Trainees in General Surgery. Toale C, Morris M, Roche A, Voborsky M, Traynor OJ, Kavanagh DO

- Oral Presentation, Association of Surgeons of Great Britain and Ireland International Surgical Conference, Harrogate, United Kingdom, 19/05/23

Why do residents fail simulation-based assessments of operative competence? Toale C, Morris M, Kavanagh DO

- Oral presentation, Association for Surgical Education annual meeting, San Diego USA, 10/04/23

Applicant Factors Associated with Success in Surgical Training: Do Psychomotor Assessments Have A Role? Toale C, Morris M, Ryan D, Doherty E, Boland F, Traynor O, Kavanagh DO

- Plenary presentation, Sir Peter Freyer Surgical Symposium, 02/09/22
- Oral presentation, Association of Surgeons of Great Britain and Ireland, Liverpool 03/05/22

Generating A Prioritised List of Operative Procedures for Simulation-Based Assessment Using A Modified Copenhagen Needs Assessment Framework Toale C, Morris M, Roche A, Heskin L, Kavanagh DO

- Oral presentation, Sir Peter Freyer Surgical Symposium, 02/09/22

Constructing an Initial Validity Argument for a Novel Simulation-Based Assessment of Operative Competence in Higher Specialist General Surgery Training Toale C, Morris M, Kavanagh DO

- Presented at the University Hospital Waterford Virtual Research Meeting, 28/01/22

Defining operative experience targets in surgical training: A systematic review Toale C, O'Byrne A, Morris M, Kavanagh DO.

- Presented at the Sir Peter Freyer Symposium, National University of Ireland, Galway, 06/09/21 (Virtual). Winner, session prize.

Exploring The Reliability Of Workplace-Based Operative Competency Assessments In Surgical Training Using Generalisability Theory C Toale, DM Nally, DM Ryan, D O'Keeffe, D Kavanagh

- Presented at the Sir Peter Freyer Symposium, National University of Ireland, Galway, 06/09/21 (Virtual)
- Presented at the Association of Surgeons of Great Britain and Ireland, Liverpool 03/05/22

A Novel Simulation-Based Assessment of Operative Competence in General Surgery Higher Specialist Training: A Mixed-Methods Study

- Presented at the Sir Peter Freyer Symposium, National University of Ireland, Galway, 04/09/23

## List of published scientific papers for consideration in this thesis

*Papers meeting the requirements of PhD by publication*

1. Toale C, Morris M, Kavanagh D. Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review. Am J Surg. 2021 Oct;222(4):723-730. doi: 10.1016/j.amjsurg.2021.03.035. Epub 2021 Apr 6. PMID: 33849711.

- Authorship status: First
- Article Type: Review (Systematic, Scoping)
- Journal: American Journal of Surgery. Q1 (Surgery, Medicine [miscellaneous]). Impact Factor 3.125 (2021)

2. Toale C, Nally DM, Ryan DM, Morris M, Kavanagh DO. How do Trainers and Trainees use Formative Workplace-based Assessments of Operative Competence to Deliver and Receive Structured Feedback? J Surg Educ. 2022 Mar-Apr;79(2):485-491. doi: 10.1016/j.jsurg.2021.08.023. Epub 2021 Sep 28. PMID: 34593328.

- Authorship: First
- Article Type: Original Research
- Journal: Journal of Surgical Education, Q1 (Surgery, Education). Impact factor 3.524 (2021)

3. Toale C, Morris M, Kavanagh DO. Assessing Operative Skill in the Competency-based Education Era: Lessons From the UK and Ireland. Ann Surg. 2022 Apr 1;275(4):e615-e625. doi: 10.1097/SLA.0000000000005242. PMID: 35129531.

- Authorship status: First
- Article Type: Review (Systematic, Scoping)
- Journal: Annals of Surgery, Q1 (Surgery). Impact Factor 13.79 (2021)

4. Toale C, Morris M, Ryan DM, Boland F, Doherty EM, Traynor OJ, Kavanagh DO. Baseline Assessments of Psychomotor Aptitude are Associated with the Future Operative Performance of Surgical Trainees A Prospective Observational Cohort

Study. Ann Surg. 2022 Jul 15. doi: 10.1097/SLA.0000000000005565. Epub ahead of print. PMID: 35837886.

- Authorship status: First
- Article Type: Original Research
- Journal: Annals of Surgery. Q1 (Surgery). Impact Factor 13.79 (2021)

5. Toale C, O'Byrne A, Morris M, Kavanagh DO. Defining operative experience targets in surgical training: A systematic review. Surgery. 2022 Nov;172(5):1364-1372. doi: 10.1016/j.surg.2022.07.015. Epub 2022 Aug 26. PMID: 36038374.

- Authorship status: First
- Article Type: Review (Systematic)
- Journal: Surgery. Q1 (Surgery). Impact factor 4.348 (2021)

6. Toale C, Morris M, Kavanagh DO. Training and assessment using the LapSim laparoscopic simulator: a scoping review of validity evidence. Surg Endosc. 2022 Sep 19. doi: 10.1007/s00464-022-09593-0. Epub ahead of print. PMID: 36123545.

- Authorship: First
- Article Type: Review (Systematic)
- Journal: Surgical Endoscopy and other Interventional Techniques. Q1 (Surgery). Impact Factor 3.453

7. Toale C, Morris M, Kavanagh DO. Perspectives on simulation-based assessment of operative skill in surgical training. Med Teach. 2022 Oct 26:1-8. doi: 10.1080/0142159X.2022.2134001. Epub ahead of print. PMID: 36288727.

- Authorship: First.
- Article Type: Original Research
- Journal: Medical Teacher. Q1 (Education, Medicine [miscellaneous]). Impact Factor 4.277

8. Toale C, Morris M, O'Keeffe D, Boland F, Ryan DM, Nally DM, Kavanagh DO. Assessing operative competence in core surgical training: A reliability analysis. Am J

Surg. 2023 Jun 30:S0002-9610(23)00273-8. doi: 10.1016/j.amjsurg.2023.06.020. Epub ahead of print. PMID: 37481408.

- Authorship: First
- Article Type: Original Research
- Journal: American Journal of Surgery. Q1 (Surgery). Impact Factor: 2.7

9. Toale C, Morris M, Konge L, Nayahangan LJ, Roche A, Heskin L, Kavanagh DO. Generating a prioritised list of operative procedures for simulation-based assessment of General Surgery Trainees through consensus

- Authorship: First
- Article Type: Original Research
- Journal: Accepted for publication, Annals of Surgery. Q1 (surgery). Impact Factor 13.79 (2021)

**Related published scientific papers**

1. Toale C, Morris M, Kavanagh DO. Training to proficiency in surgery using simulation: is there a moral obligation? J Med Ethics. 2022 Jan 6:medethics-2021-107678. doi: 10.1136/medethics-2021-107678. Epub ahead of print. PMID: 34992083.

- Authorship: First
- Article type: Opinion/ Editorial
- Journal: Journal of Medical Ethics, Q1 (Arts and Humanities [miscellaneous], Health Policy, Health [social science], Issues, Ethics and Legal Aspects. Impact Factor: 5.926 (2021)

2. Toale C, Morris M, Kavanagh DO. Comment on: The impact of virtual-reality simulation training on operative performance in laparoscopic cholecystectomy: meta-analysis of randomized clinical trials. BJS Open. 2022 Nov 2;6(6):zrac134. doi: 10.1093/bjsopen/zrac134. PMID: 36412113.

- Authorship: First
- Article Type: Correspondence
- Journal: British Journal of Surgery Open. Q1 (Surgery, Medicine [miscellaneous]). Impact Factor: 3.875.

3. Toale C, O'Byrne A, Morris M, Kavanagh DO. Characterizing individual trainee learning curves in surgical training: Challenges and opportunities. Surgeon. 2022. doi: 10.1016/j.surge.2022.11.003

- Authorship: First
- Article type: Editorial
- Journal: The Surgeon. Q2 (Surgery). Impact Factor 2.632

4. Toale C, Morris M, Kavanagh DO. Why do residents fail simulation-based assessments of operative competence? A qualitative analysis. *Global Surg Educ* **2**, 83 (2023). https://doi.org/10.1007/s44186-023-00161-1

- Authorship: First

- Article Type: Original Research
- Journal: Global Surgical Education – Journal of the Association for Surgical Education. Impact factor – not listed.

**Related unpublished supporting scientific papers**

1. Toale C, Morris M, Gross S, O'Keeffe D, Ryan DM, Boland F, Doherty E, Traynor OJ, Kavanagh DO. Performance at National Selection and Future Performance in Surgical Training

- Authorship: First
- Article Type: Original Research
- Status: Under Review, JAMA Surgery

2. Toale C, Morris M, Roche A, Voborsky M, Traynor O, Kavanagh DO. Development and Validation of a Simulation-Based Assessment of Operative Competence for Higher Specialist Trainees in General Surgery.

- Authorship: First
- Article Type: Original Research
- Status: Under review, British Journal of Surgery

3. Toale C, Morris M, Julius B, Strawbridge J, Kavanagh DO. Using simulation to assess operative competence: a qualitative case study of a novel simulation-based assessment

- Authorship: First
- Article Type: Original Research
- Status: Final Draft

# Chapter 1. Introduction: The current status of high-stakes technical performance assessments in surgical training

## Chapter overview

This chapter outlines the current status of simulation as a high-stakes assessment modality in surgical training, introducing the current research in this area. This chapter takes the form of a review article. Mr. Toale is the first author of this work he conducted the literature search, synthesised the findings, and drafted the initial manuscript.

## Abstract

### Purpose

Simulation as a surgical training modality is well established, though its use as in assessment is less well explored. In particular, the use of simulation-based assessment (SBA) in high stakes assessment such as recruitment, entrustment, and certification, is employed to varying degrees across jurisdictions. The aim of this review is to collate and critique the published evidence regarding the utility of SBA in high stakes assessments of operative skill.

### Methods

The search strategy for this review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines 2020. The Cochrane library, PubMed (Medline), Embase and Web of Science databases were searched for all articles reporting utility evidence for the use of technical performance assessments in postgraduate surgical training. Included articles were those reporting on the use or proposed use of such assessments in high-stakes scenarios, namely selection, progression through training, entrustment or autonomy granting, credentialing or certification.

### Results:

Sixty-nine articles were included in a narrative synthesis. Performance assessments using simulation or psychomotor skill evaluations have been employed across jurisdictions for selection of surgical trainees, with a growing evidence base regarding their predictive validity. While a number of studies have demonstrated the benefit of SBA prior to entrustment or autonomy granting, no studies have reported on the implementation or consequence validity of such an approach to progressive autonomy granting. There are a few notable examples of SBA at end-of-training certification, though implementation varies across specialties and jurisdictions.

**Conclusions:**

In spite of evidence demonstrating the validity of simulation as an assessment modality in surgical training, there remains an implementation gap with respect to its use in high-stakes settings. Future studies should explore the barriers to implementation of SBA in trainee selection, entrustment, progression-through-training decisions, and ultimate certification.

## 1.1 Introduction

Simulation is an established modality in surgical training, though its use in assessment is less well explored. As evidence supporting the validity of simulation-based assessment (SBA) continues to grow, its potential for use in high-stakes assessment will come under scrutiny. High-stakes assessments are those used to guide decisions regarding selection, progression and certification[19]. Within the context of procedural training, entrustment or autonomy-granting is another form of high-stakes assessment[20]. Robust evidence of validity, feasibility, cost-effectiveness and acceptability are of primary concern in this context[19].

While curricula emphasise longitudinal workplace-based assessment in surgical training, high stakes assessments of competence remain important in maintaining standards and public trust[19]. Certification exams such as the Fellowship of the Royal College of Surgeons (FRCS) examination or American Board of Surgery (ABS) examinations are primarily knowledge based, though assessment of operative skill has been incorporated into some processes, such as the Fellow of the European Board of Vascular Surgery (FEBVS) examination.

SBA has potential applicability beyond end-of-training certification. Technical aptitude assessments could be used in trainee selection processes. Performance in SBA could inform decisions regarding progression through training. Passing a high-stakes simulation-based assessment or proficiency benchmarked training and assessment program could be deemed a pre-requisite prior to in-theatre operating, or could inform progressive entrustment decisions. The aim of this review is to explore the current status of SBA in high-stakes technical skill assessment in training.

## 1.2 Methods

This review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement 2020[21]. Search terms included Medical Subject Headings (MeSH) and free-text terms; (Simulation) AND (Certification OR Selection OR Recruitment OR credentialing OR Progression OR entrustment OR privileging OR autonomy granting) AND (Surgery) AND (internship and residency OR training OR residency education) AND (Assessment). The Cochrane library, PubMed (MedLine), Embase and Web of Science databases were searched (last search May 2023). The full search strategy is outlined in Appendix 1.1, supplemental data.

After screening for duplicates, titles and abstracts were screened for relevance according to the inclusion criteria outlined in Table 1.1. Performance assessments were defined as assessments taking place outside of the operating room environment, to include simulation-based skill assessments and technical aptitude testing. Data were sought from all studies reporting on aspects of utility relating to high-stakes technical performance assessments in surgery, namely; validity, reliability, acceptability, cost or educational impact[22]. 'High-stakes' applications in surgery included trainee selection or recruitment, progression through training, credentialing, or certification. For the purposes of this manuscript, certification is defined as the process whereby the qualifications of a healthcare practitioner are assessed and verified[23]; in the context of surgical training this can be thought of as the certification of a surgical trainee as an independent practitioner. Credentialing is the process whereby 'a specific scope and content of patient care services are authorized for a practitioner'[23]. In surgical training we can relate this to the assessment of procedure-specific or modality-specific competencies.

*Table 1.1: Inclusion and Exclusion Criteria – The current status The current status of high-stakes technical performance assessments in surgical training*

| **Inclusion criteria** | **Exclusion criteria** |
|---|---|
| Experimental and quasi-experimental studies | Editorials, opinion pieces, case studies or technique articles |
| Studies exploring the utility of performance assessments during, or on application to, surgical training (including at fellowship level) | Studies exploring the utility of performance assessments to inform high-stakes assessments in medical school only. |
| Studies pertaining to all surgical specialties and sub-specialties, including gynaecological and ophthalmological surgery | Studies pertaining to non-surgical specialties (e.g. interventional procedures in cardiology or gastroenterology) |
| Studies exploring the use of high-stakes assessments in determining surgical or operative skill | Studies assessing skill or competence in endoscopic or bedside procedures. |
| | Studies assessing non-technical skills. |
| Studies pertaining to assessments designed to inform high-stakes decisions. | Studies exploring simulation as a training modality only. Studies pertaining to formative or low-stakes assessment. |
| Full text available | Conference abstracts |

The quality of included studies was assessed using the Medical Education Research Study Quality Instrument (MERSQI), a 10-item quality assessment tool scored to a maximum of 18[24]. Due to the anticipated heterogeneity of included studies, findings were synthesised through narrative review.

### 1.3 Results

A total of 2,319 records were retrieved from database searching, of which 580 were duplicates. On screening, 1,575 records were excluded, with 164 reports sought for retrieval. Of these, 12 could not be retrieved (published abstract only). Of the 152 full-text articles assessed for eligibility, 69 were ultimately included in this review (Figure 1.1), with findings presented using narrative synthesis.

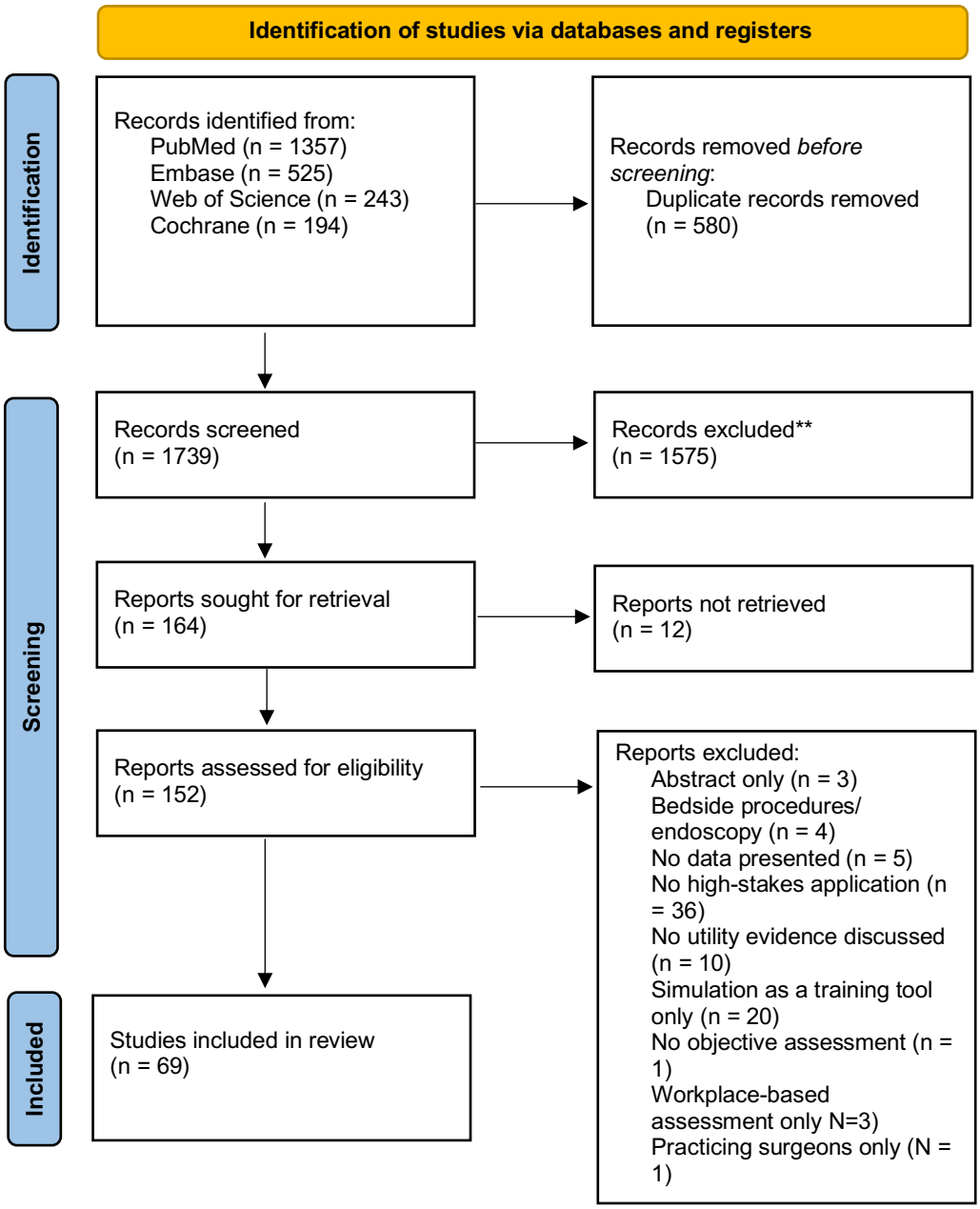



*Figure 1.1: PRISMA flow-chart, The current status of simulation as a high-stakes technical skill assessment modality in surgical training From:* Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

The Medical Research Study Quality Instrument (MERSQI) score of included studies ranged from 6.5 – 16 (Appendix 1.2, supplemental data). Study characteristics are outlined in Table 1.2.

*Table 1.2: Characteristics of included studies – the current status of simulation as a high-stakes technical skill assessment modality in surgical training*

| Author | Year | Country | Participants | N | Assessment modality | High stakes application |
|---|---|---|---|---|---|---|
| Haidari | 2022 | Denmark | Novices, intermediate and expert surgeons | 41 | LapSim (Surgical Science) | Credentialing |
| Gangemi | 2022 | USA | General surgery residency applicants and medical students | 32 | Manual dexterity (MD) was assessed through six tasks: O'Connor Tweezer Dexterity Test and Minnesota Manual Dexterity Test; Peg Transfer Test Fundamentals of Laparoscopic Surgery (box); Ring and Rail, Thread the Ring and Suture Sponge (da Vinci Surgical Simulator). | Selection |
| Lund | 2022 | USA | General surgery interns | 19 | Simulation | Selection |
| Fong | 2022 | United Kingdom | Executive council of the Association of Laparoscopic Surgeons of Great Britain and Ireland (ALSGBI) | 76 | LapPass, laparoscopic box trainer | Credentialing |
| Toale | 2022 | Ireland | representatives from identified key stakeholder groups; executive management, risk management, a practicing surgeon, an anaesthesiologist, a theatre-nursing representative, a representative from simulation industry, a patient, a medical student, a junior surgical trainee, and a senior surgical trainee. | 10 | Simulation (not further specified) | Selection, credentialing (autonomy granting), certification |
| Engberg | 2021 | Denmark | Novices, anaesthesiologists and endovascular experts | 42 | Synthetic model (mannikin), Resuscitative Balloon Occlusion of the Aorta (REBOA) | Certification |
| Seaman | 2021 | USA | General surgery residents and residents in Obstetrics and Gynaecology | 176 | Simulation (FLS) | Certification/ credentialing |
| Doneza | 2021 | USA | Obstetricians and gynaecologists | 108 | Simulation | Credentialing (maintenance) |
| Hermsen | 2021 | USA | Cardiothoracic fellows (PGY 6 - 8) and attending surgeons | 6 | Simulation | Certification |
| Karani | 2021 | USA | Applicants to a urology surgery residency | 94 | Benchtop simulators | Selection (residency) |
| Aljamal | 2020 | USA | General surgery residents | 44 | Multi-station simulation-based OSCE | Credentialing (entrustability) |
| Goderstad | 2020 | Norway | Gynaecology trainees | 12 | LapMentor (Symbionix) | Credentialing |
| AlJamal | 2020 | USA | Interviewees to a general surgery residency | 60 | Simulation | Selection |
| Gillis | 2020 | Canada | Postgraduate year 3-5 trainees of a 5-year Canadian orthopaedic residency program | 34 | Simulation-based OSCE, cadaveric (procedural simulations) | Certification |
| Satava | 2020 | USA | Multi-specialty subject-matter-experts in robotic surgery | 80 | Robotic simulation | Credentialing |
| Nayahangan | 2019 | Denmark | Novices and experienced open vascular surgeons | 25 | Synthetic abdominal model for open aortic surgery, Vascular International School AG, Kerns, Switzerland) | Certification |
| Halwani | 2019 | Canada | General surgery residents | 19 | Simulation | Certification |
| Bube | 2019 | Denmark | Urology doctors (novices, intermediate and expert) | 49 | TURB Mentor VR simulator (3D Systems/Simbionix) | Credentialing (autonomy granting) |
| Sousa | 2019 | Portugal | Candidates undertaking the Fellow of the European Board of Vascular Surgery (FEBVS) examination | 300 | multi-station open technical skills assessment (aortic anastomosis, distal anastomosis, carotid endarterectomy) and endovascular technical skills assessment | Certification |
| Bilgic | 2018 | Canada | Fundamentals of Laparoscopic Surgery (FLS) candidates | 7232 | FLS exam using a box trainer | Credentialing |
| Hovgaard | 2018 | Denmark | Novice gynaecological surgeons and 11 experienced gynaecological robotic surgeons | 22 | RobotiX Mentor (3D Systems, Colorado, USA) | Credentialing |
| Erturan | 2018 | United Kingdom | Orthopaedic consultant surgeons, trainees and novices (interns and medical students) | 52 | a standardized bench-top simulated hip arthroscopy task | Selection (fellowship) |
| de Vries | 2018 | The Netherlands | Urologists and urology residents | 76 | The Simbla TURBT simulator (SAMED GmbH, Dresden, Germany) | Certification |
| Moglia | 2018 | Italy | Medical students without prior simulator experience | 166 | dV-Trainer (Mimic, Seattle, WA,USA) | Selection |
| de Montbrun | 2018 | Canada | General surgery program directors | 14 | Simulation (not further specified) | Certification |
| Ahmed | 2018 | International panel | Experts from EULIS, EUREP, ESU and ESUT | 35 | Simulation (not further specified) | Autonomy granting |
| Masood | 2018 | USA | Applicants to an ENT residency program | 41 | Six simulation modules reflecting tasks used in six subspecialties within otolaryngology | Selection |
| Awad | 2018 | United Kingdom | International experts in advanced minimally invasive surgery | 57 | Simulation (not otherwise specified) | Credentialing |

| | | | | | | |
|---|---|---|---|---|---|---|
| Louridas | 2017 | Canada | Pre-clinical medical students | 65 | Laparoscopic and open technical tasks: peg transfer, circle cutting, intracorporal knot tie, one handed tie and simulated laparotomy closure | Selection |
| Hazboun | 2017 | USA | Prospective paediatric surgery fellowship candidates | 40 | Simulated laparoscopic intracorporal knot tying | Selection (fellowship) |
| Gardner | 2017 | USA | Surgery faculty/ residents/ staff and medical students | 43 | "Sim Games" including bench models for basic knot tying and basic suturing | Selection |
| Dwyer | 2016 | Canada | PGY-1 and PGY-4 orthopaedic surgery residents | 18 | Synthetic models | Credentialing (entrustability) |
| Winkler-Schwartz | 2016 | Canada | Applicants to a neurosurgery residency programme | 22 | NeuroTouch virtual reality simulator | Selection (residency) |
| Thinggaard | 2016 | Denmark | Novices and experienced laparoscopic surgeons | 40 | TABLT training and testing tool using a portable laparoscopic trainer | Credentialing (autonomy granting) |
| Ritter | 2016 | USA | An 'executive group' consisting of 5 members in leadership positions in the Fundamentals of Laparoscopic Surgery (FLS) proctor program, and a 'subject matter group; of experienced proctors | 20 | Fundamentals of Laparoscopic Surgery (FLS), laparoscopic box trainer | Credentialing |
| Louridas | 2016 | Canada | International educational directorates | 8 | Aptitude tests, simulation | Selection, progression through training, certification |
| Janssens | 2015 | Australia | Gynaecology registrars | 16 | LapMentorII(Simbionix,Cleveland,OH,USA) | Credentialing |
| Jacobsen | 2015 | Demark | Experienced arthroscopic surgeons and novices | 26 | The ARTHRO Mentor virtual reality knee arthroscopy simulator | Autonomy granting |
| Duran | 2015 | USA | Residents, fellows, and surgeons with extensive endovascular training | 20 | Silicon model | Credentialing |
| Thomsen | 2015 | Denmark | Ophthalmology trainees, cataract surgeons and vitreoretinal surgeons | 42 | EyeSi™ simulator (Haag-Streit Simulation) | Certification/ credentialing |
| Groenier | 2015 | The Netherlands | Novices | 98 | Visual-spatial ability testing the mental rotation test, the Paper folding test, the surface development test and the Rotating shapes test. Spatial memory tests; The Corsi block tapping test. Perceptual speed; Number comparison test and the Identical pictures test. Reasoning; The Raven advanced progressive matrices test and a verbal reasoning test from the Groninger intelligence test | Selection |
| Rooney | 2015 | USA | Registered Fundamentals of Laparoscopic Surgery (FLS) proctors | 42 | FLS exam using a box trainer | Credentialing |
| Ahmed | 2015 | International panel | Attendants at European Association of Urology, and robotic Urology Section meetings | Not described | robotic simulation | Credentialing |
| Deutsch | 2015 | USA | Otolaryngology program directors | 43 | Simulation (not otherwise specified) | Credentialing |
| Culligan | 2014 | USA | Robotic surgery experts and credentialed gynaecology surgeons naïve to robotic surgery | 19 | Porcine laboratory, automated performance metrics measured by the DaVinci platform | Credentialing (robotic) |
| Gallagher | 2014 | Ireland | Applicants to specialty training in general surgery | 98 | Multi-station OSCE across ten skills stations | Selection (Progression through training) |
| Buckley | 2014 | Ireland | Medical students without prior simulator experience | 20 | Visual-spatial ability, depth perception and psychomotor ability | Selection |
| Giannotti | 2014 | Italy | General and bariatric surgeons | 20 | Lap Mentor simulator | Certification |
| Moglia | 2014 | Italy | Medical students | 121 | da Vinci Skills simulator | Selection |
| Burden | 2014 | United Kingdom | National experts in gynaecological laparoscopy and gynaecology trainees | 37 | Simulation (not otherwise specified) | Credentialing (autonomy granting) |
| Buckley | 2013 | Ireland | Medical students | 80 | Visual-spatial ability, depth perception and psychomotor ability | Selection |
| Duran | 2013 | USA | Vascular surgery trainees | 183 | Fundamentals of Endovascular Surgery (FEVS) – silicon and virtual reality models | Certification |
| Rooney | 2012 | USA | Surgeons and non-surgeons attending Fundamentals of Laparoscopic Surgery proctor training workshops | 30 | Fundamentals of Laparoscopic Surgery (FLS), laparoscopic box trainer | Credentialing |
| Stefanidis | 2012 | USA | Members of the Association for Surgical Education | 60 | Simulation (not otherwise specified) | Credentialing |
| Alvand | 2011 | United Kingdom | Medical students | 33 | Benchtop simulators | Selection |
| Panait | 2011 | USA | Applicants to a community general surgery program | 42 | LapSim (Surgical Science) | Selection |
| Van Herzeele | 2010 | United Kingdom | Medical students | 20 | Fine motor dexterity (Perdue pegboard), psychomotor ability (minimally invasive virtual reality surgical trainer, MIST-VR), image recall (Rey-Osterrieth) and organisational aptitude (map-planning) | Selection |

| | | | | | | |
|---|---|---|---|---|---|---|
| Xeroulis | 2009 | Canada | Junior residents (PGY 1 - 3), Senior residents (PGY 4,5) and staff surgeons | 26 | Fundamentals of Laparoscopic Surgery (FLS), laparoscopic box trainer | Credentialing |
| Carroll | 2009 | Ireland | Applicants to specialty training in Plastic surgery | 34 | Synthetic models of four plastic surgery procedures | Selection (specialty training) |
| Salgado | 2009 | USA | Applicants to a Minimally Invasive Surgery (MIS) fellowship | 8 | | |
| Gallagher | 2008 | Ireland | Applicants to higher specialist training in general surgery | 16 | Six open stations used high-fidelity realistic bench models(Limbs & Things, Bristol, UK and AnnexArt, Anglesey, Wales). Four endoscopic surgery stations using the ProMis laparoscopy simulator (Haptica, Dublin, Ireland) and the GI Mentor endoscopy simulator (Simbionix, Cleveland, OH, USA). | Selection, progression through training |
| Goldberg | 2008 | USA | Fourth year medical students | 113 | Stromberg Dexterity Test of gross motor function and the O'Conner Tweezer Test of fine motor dexterity | Selection |
| Tedesco | 2008 | USA | Vascular surgery | 17 | Procedicus Vascular Intervention System Trainer (VIST) endovascular simulator (Mentice Corporation, Gothenburg, Sweden) | Selection (fellowship), credentialing |
| Cope | 2008 | Australia | Interns in their first post-graduate year | 22 | MIST VR simulator (Mentic, Sweden) | Selection |
| Tansley | 2007 | United Kingdom | Applicants to specialty training | 54 | Visuospatial and technical ability testing | Selection (specialty training) |
| Pandey | 2006 | International (European Board of Vascular Surgery | Candidates and examiners in a pilot examination of technical competence | 30 | Assessment of technical competence examination, synthetic models | Certification |
| Aggarwal | 2006 | United Kingdom | Expert and novice surgeons with respect to cholecystectomy volume | 40 | LapSim (Surgical Science, Gothenburg, Sweden) | Credentialing (autonomy granting) |

### *1.3.1 Resident selection or recruitment:*

#### *1.3.1.1 Psychomotor aptitude assessments:*

Studies have demonstrated the validity of aptitude assessments as a resident selection modality. These assessments can stratify candidates according to ability[25], and display predictive validity: Buckley et al. (2014) describe the use of visuo-spatial, depth perception and dexterity assessments to identify students with 'high aptitude', who achieved proficiency in simulated laparoscopic suturing at a faster rate than those with 'low aptitude'[26]. Thirty percent failed to reach proficiency with repeated attempts[26], with similar findings reported by other studies[27 28]. Recent work from our institution suggests that aptitude assessments are associated with future operative performance[29]. These assessments may complement other selection criteria; in studies by Gangemi et al. (2022) and Goldberg et al. (2008), manual dexterity scores did not correlate with United States Medical Licensing Exam (USMLE) results or medical school academic grades[27 30], suggesting that they assess a distinct construct.

#### *1.3.1.2 Simulated operative tasks:*

Simulated basic surgical tasks have been used to stratify candidates based on ability at the time of selection[31 32]. Such assessments can identify candidates who struggle to reach task proficiency[33], and results appear not to be associated with self-assessed aptitude[34 35]. Lund et al. (2022) report that SBA scores at selection were associated with higher future milestone assessment scores ($\beta = .45$, $p = 0.03$) and higher American Board of Surgery In-Training Examination (ABSITE) scores ($\beta = .43$, $p = 0.02$), while faculty interview scores and United States Medical Licensing Examination (USMLE) scores were not[36].

#### *1.3.1.3 Implementation and feasibility*

In a survey study of international educational directorates by Louridas et al. (2016), technical assessments informed selection in Australia, Ireland, New Zealand and the United Kingdom, but not in other jurisdictions surveyed (Canada, Hong Kong, Sweden)[37]. Simulated operative task assessments were incorporated into a general surgery residency selection programme by Al Jamal et al.(2020) at the Mayo Clinic (Rochester, Minnesota, USA). Resource implications are discussed; the multi-station assessment took 192 minutes to administer to 60 applicants[38], demonstrating feasibility. Baseline aptitude assessments have been incorporated into the Royal College of Surgeons in Ireland's (RCSI) trainee selection process[26], though further examples of such selection frameworks are lacking.

#### *1.3.1.4 Acceptability*

In Al Jamal's (2020) study, the incorporation of SBA stations did not negatively impact the candidate's overall perception of an interview process[38]. However, while some candidates reported that the experience was 'great' and 'engaging', others described it as 'intense', 'too stressful', and 'too high pressure'[38]. Masood et al. (2018) explored the potential use of simulated assessments at selection into an otolaryngology residency program; 98.5% of participants strongly agreed that the simulations were enjoyable, and 92.7% recommended their use in future selection, while 63.5% strongly disagreed that the simulations were stressful[39]. Importantly, the contribution of scores to selection decisions was not discussed in this study[39]. Rather, feasibility and acceptability of such an approach was explored without using SBA-derived scores to contribute to ranking.

In Louridas et al.'s 2016 study, four of eight educational directorates surveyed agreed or strongly agreed that technical performance should be assessed during selection, while 50% were 'neutral' on the question[37]. Lack of validity evidence was reported as a barrier to its use[37]. In a qualitative study of surgical education stakeholders at our institution, interviewees believed that assessing technical skill as

part of selection was fair as part of a multi-factorial process, though some participants believed that candidates should not be barred from entry based on aptitude or 'innate ability' alone[40]. Gardner et al. (2017) describe the acceptability of knot-tying and suturing assessments as part of resident selection[41]. Faculty, residents, staff and students believed that assessing *non-technical* skills was more valuable than technical skill assessment during the selection process[41]. Concerns were raised regarding time and resources required to implement SBA at selection[41]. Technical skills assessed were believed to be those that could be easily taught during residency, though it was believed that SBA could assess *non-technical* skills such as learning *aptitude* and receptivity to feedback[41].

### *1.3.2 Progression through training and remediation:*

#### *1.3.2.1 Progression to the next training year:*

Gillis et al. (2020) describe the validation of the Surgical OSCE (S-OSCE), designed to inform trainee advancement decisions in orthopaedic residency[42], mapped to the Canadian Medical Education Directives for Specialists (CanMEDS) framework. Moderate correlations between scores awarded and training year were observed ($r^2 = 0.5737$). Performance strongly correlated with entrustability scores (O-Scores) ($r = 0.89$). A feasibility analysis showed the S-OSCE to be financially practicable, with total costs calculated at $16,044.32[42] per session.

#### *1.3.2.2 Progression from 'Core Training' to 'Higher Specialist' training*

The Royal College of Surgeons in Ireland (RCSI) have described the implementation of SBA to inform decisions on progression from core surgical training (CST) to higher specialist training (HST). Gallagher et al. (2008)[43] describe a 10-station (20-30 minutes per station) skills assessment, scored by trained examiners. Candidates who were ultimately selected performed better in technical assessments, and a strong correlation was observed between SBA scored and overall multifaceted progression scores (incorporating scores for educational and academic record,

record in CST assessments, and research output ) ($r = 0.746$, $p<0.001$)[43]. A similar study was conducted with applicants to a plastic surgery HST programme; over 6 stations, the candidates ultimately selected performed 2.2 times better than those who were not ($p<0.01$)[44]. SBA improved the statistical power of the multi-factorial selection process[44]. Further validity evidence was described in 2014; candidates with higher technical skill assessment scores were more likely to progress to HST, and skill assessment scores were moderately correlated with the total score awarded[45]. The ultimate selection outcome was most strongly predicted by the technical skill assessment scores on regression analysis[45]. Evidence of SBA use at this progression point outside of Ireland is lacking. Nine HST programmes in London were surveyed by Tansley et al. (2007); at this time, one general surgery department had piloted testing of visuospatial and technical ability at this progression stage[46].

*1.3.2.3 Progression to fellowship training*

Hazboun et al. (2017) describe the implementation of SBA into the selection process for a paediatric fellowship programme[47]. Candidates were observed performing a simulated laparoscopic intra-corporeal knot. Scores awarded significantly altered the ranking of candidates otherwise obtained through Electronic Residency Application Service (ERAS) application and interview. Validity evidence, with respect to the relationship between scores awarded and other variables, is retrospectively outlined; no correlation was identified between SBA results and passing the Fundamentals of Laparoscopic Surgery (FLS) exam[47]. Erturan et al. (2018) used an arthroscopy simulator to assess candidate proficiency, to inform selection for a hip arthroplasty fellowship[48]. SBA scores correlated with arthroplasty experience[48]. Salgado et al. (2009) used the LapSim™ (Surgical Science, Sweden) simulator to assess the technical skill of candidates applying to a minimally invasive surgery fellowship[49]. There was no significant correlation between assessment metrics and interview scores[49], implying that SBA assesses a distinct construct compared to traditional interview. In a 2012 cross-sectional survey of US fellowship program directors, none reported using SBA at the time of selection[50].

*1.3.2.4: Acceptability of SBA used to inform progression through training*

Seven of eight educational directorates surveyed by Louridas et al (2016) assessed technical skill during training using a combination of workplace-based assessment and SBA[37]. Operating room assessments were preferred over those in the simulation laboratory[37]. Summative technical skill assessments used to inform progression to the next training year were reported as 'less appealing'[37]. Financial limitations were a barrier to technical skill assessment during training[37]. In a qualitative study of key stakeholders at our institution, participants felt that the aim of SBA during training should be to identify candidates in need of extra support, and should only be used to halt progression in cases of repeated failures or gross incompetence[40]. It may be more acceptable to mandate a basic level of competence to be assessed using SBA in early training; in a survey-based study of gynaecological surgery experts and trainees, 77.7% of experts and 64.7% of trainees agreed or strongly agreed that junior trainees should pass a basic laparoscopic skills curriculum prior to progressing to the next training year[51].

*1.3.3 Entrustment, Credentialing, and procedure-specific certification*

*1.3.3.1 Validity and Implementation*

A number of studies report validity evidence for SBA designed for use in entrustment or procedure-specific credentialing[52-64]; Further studies report on the development of SBA for credentialing in particular modalities, for example generic laparoscopic[65] or endovascular[66] skills. Culligan et al. (2014) described a proficiency-based progression in robotic gynaecological surgery[67]. Fourteen novice surgeons trained to pre-defined benchmarks on simulated tasks prior to undergoing animal-model based training, and finally their first human robotic supracervical hysterectomy[67]. This led to similar operative times to expert robotic surgeons[67]. In spite of studies outlining validity arguments for the use of simulation in this context, articles reporting the implementation of such programmes are lacking. In a survey of US otolaryngology program directors, 91% of 34 programmes were not using simulation for

credentialing purposes, while a further two programs (6%) reported doing so to a limited extent, with only one program using it for this purpose extensively[68]. In Janssen et al.'s 2015 study, sim-based credentialing in laparoscopy did not result in increased autonomy granting for trainees[69].

*1.3.3.2 Acceptability*

In a study by Burden et al. (2014), 74% of experts and 35.3% of trainees agreed or strongly agreed that passing a basic laparoscopic skills curriculum should be a prerequisite for in-theatre operating[51]. A simulated hand-eye co-ordination assessment was reported to be the most appropriate assessment modality in this context[51]. Further views of robotic surgery experts are explored in a consensus study by Ahmed et al. (2015): participants agreed that trainees must undergo simulation-based training before in-theatre operating[70]. Conversely, participants in a Delphi study by Awad et a. (2014) could not reach consensus on whether simulation-based assessment should be used for advanced minimally-invasive surgical skill evaluation[71].

*1.3.4 Certification:*

*1.3.4.1 Mandatory in-training assessments*

A number of studies outline utility evidence for the fundamentals of laparoscopic surgery (FLS) examination, an assessment of manual and cognitive laparoscopic skills with a pass score required for certification in a number of surgical specialties in the USA[72-76]. The LapPass assessment (Fong et al. 2021) has been similarly designed for the UK context[77], though is not yet a mandatory component of UK surgical training. The Fundamentals of Endovascular Surgery (FEVS) exam has also been designed. Despite documented validity evidence such an exam has not been incorporated into US-based vascular training[78 79]. Satava et al. and colleagues (2020 – 2021) outline their approach to the development of a Fundamentals of Robotic Surgery (FRS) curriculum, to assess skill prior to progression to clinical surgery[80-82];

it has yet to be formally incorporated into national certification or credentialing processes in robotic surgery.

*1.3.4.2 End of training certification*

The European Board of Surgery Qualification in Vascular Surgery (EBSQVS) examination incorporates an assessment of candidates' technical skill, and is required for certification in some jurisdictions[83]. Paney et al. (2006) describe the validity evidence from 22 candidates and 8 examiners; construct validity (assessor-candidate comparisons), inter-observer reliability (α = 0·91), and inter-station reliability ((rs) = 0·79, P < 0·001) data are reported[84]. No significant correlation was observed between the SBA scores and scores derived from a logbook of operative experience[84].

De Montbrun et al. (2013) described the development of the Colorectal Objective Structured Assessment of Technical Skill (COSATS), to assess trainee competence at certification level[14]. This study demonstrated acceptable reliability metrics, and could distinguish between colorectal and general surgery residents[14]. In 2016, this study group further outlined the implementation of this assessment in the American Board of Colon and Rectal Surgery (ABCRS) certification process. The COSATS exam was a mandatory component of certification in 2014, when 70 trainees were assessed. Reliability metrics were appropriately high for use in high-stakes assessment (>0.80). A low positive correlation of r = 0.25 was recorded between COSATS scores and oral examination scores[18]. Halwani et al. (2019) similarly describe the development of the General Surgery Objective Structured Assessment of Technical Skill (GOSATS), designed as an eight station simulation-based examination for graduating general surgery residents in Canada[85]. Validity evidence across the five domains of Messick's validity framework were discussed, and reliability indices appropriate for high-stakes assessment were recorded[15 86].

Acknowledging calls by key figures in cardiothoracic surgery to consider incorporating technical skill assessment in the certification process[87 88], Hermsen et

al. (2021) describe the development of a high-fidelity model for use in assessment of cardiopulmonary bypass skills[89]. Their findings assessing a small group of trainees and experts demonstrated expert-fellow differences, and may in future lead to the development of competence benchmarks for use in high-stakes assessments, though no data regarding further implementation is reported[89]. Giannotti et al. (2014) explored the use of the Lap Mentor ™ simulator as a certification tool for bariatric surgeons. They describe significant differences in simulator-measured performance metrics between general surgeons and bariatric surgeons[90], though further implementation of a simulation-based certification programme is not described.

*1.3.4.3 Acceptability*

In a 2016 study by Louridas et al., respondents reported that technical skill examinations were not assessed as a separate high-stakes examination in any jurisdiction (Australia, Canada, Denmark, Hong Kong, Ireland, the Netherlands, New Zealand, and Sweden)[37]. Assessments in the operating room were preferred over simulation-based assessments for certification purposes[37]. Lack of evidence for its use, and financial considerations, were both reported as barriers to the use of SBA at certification level[37]. In a Delphi consensus study of members of the Association for Surgical Education (Stefanidis et al., 2012), "What is the role of simulation for the certification of residents and practicing surgeons?" was identified as one of the top 10 highest-priority questions in surgical simulation[91]. Louridas et al. (2016) report that 75% of educational directorates surveyed either agreed or strongly agreed that it was important to objectively assess technical competence at the time of certification. A total of 57% of surveyed vascular trainees in the United States expect that technical skills assessment will be incorporated into the future certification process, with 52% supporting such an approach[92]. In a Delphi consensus study of 57 experts in advanced minimally invasive surgery, no agreement could be reached regarding the use of virtual reality simulation for technical skills assessment as an accreditation modality in advanced MIS; participants favoured assessment of video data in addition to key performance metrics[93].

## 1.4 Discussion

This review explored the status of SBA in operative or technical skill assessment in high-stakes scenarios in surgical training. We have outlined its use in trainee selection and informing decisions regarding trainee progression, entrustment and procedure-specific credentialing. We have further explored the evidence supporting its use in end-of training certification. The following discussion will seek to summarize where we are now, how we got here, and where we might go next.

From earlier studies demonstrating the ability of simple psychomotor assessments to predict performance in simulated tasks [25-28], our institution has implemented such assessments as part of the selection process for surgical trainees, and has demonstrated an association between aptitude scores and future performance in workplace[40]. Other institutions have provided validity evidence for the use of assessments of basic simulated operative tasks as part of the selection process[31-36] [94] [95]. While the feasibility of such an approach has been demonstrated[38], studies exploring the views of candidates and recruiters towards technical skill assessment as a selection modality outline important considerations for the future[37-41]. Studies should explore the predictive validity of these assessments beyond associations with simulated performance or performance in *early* training alone. The relationship between aptitude, skill and patient outcomes has yet to be investigated, and would be challenging without sufficiently granular data linking trainee performance data to individual operative cases.

It is the opinion of the authors that technical assessments at residency selection level or equivalent should focus on *aptitude* rather than *skill*. Knot tying, for example, represents a skill that can be mastered during training, with a low 'ceiling effect' observable in early training[96]. Furthermore, the relative experience and operative exposure of candidates at such an early stage will vary considerably; assessing for learning potential rather than amassed experience is likely more appropriate at selection. Psychomotor tests, which in theory assess underlying ability, may represent a more appropriate way to evaluate a candidate's potential over

performance in a basic surgical task. This hypothesis can and should be tested. Given that technical proficiency represents just one of many attributes expected of a competent surgeon[97], the authors believe that technical aptitude assessments should form part of a multi-factorial selection process. Such assessments could act in a complementary fashion to established academic criteria, similar to the ways in which situational judgement testing has been used to assess 'non-technical' ability[98 99]. Though a number of studies have identified that a proportion of candidates may fail to reach an acceptable technical standard even with deliberate practice[26-28], the use of psychomotor assessments alone as a screening tool for selection remains controversial[40], and may be unacceptable to applicants.

A number of robust studies have used contemporary validity frameworks to outline the development of various simulation-based assessments, intended for use in entrustment, autonomy granting or credentialing during training. The vast majority of these focus on assessing procedure-specific competence or key basic skills. More recent studies have focussed on using 'proficiency-based-progression training modules', where skills are trained to a pre-defined, usually expert-level benchmark, and formally assessed. Such an approach, incorporating simulation-based assessment, has been shown to reduce intra-operative performance errors and operative time in a recent meta-analysis[100]. This approach is perhaps best exemplified by contemporary robotic training curricula[67]. Such a model could conceivably be applied to laparoscopic, endovascular, or even open surgery, though studies reporting on the implementation of such an approach are lacking. Future studies should explore the barriers to implementing a train-to-proficiency simulation programme across surgical curricula, which would likely include significant resource and time barriers.

Further studies have described the use of SBA to inform decisions regarding progression through training. To date, these have focussed on assessing trainees at key progression points, e.g. the competitive transition from core to higher specialist training in the UK[46] and Ireland[43-45], or transition to fellowship training[47-49]. In this context, simulation is again being used as a modality to inform selection, though

appropriately should use simulated tasks or procedures to assess *learning* rather than aforementioned aptitude tests. A recent study by Gillis et al. (2020) demonstrates how using SBA in training could align with contemporary goals of competency-based education[42]. Simulation could objectively assess procedural competence during training, aligning with curriculum-derived expectations and informing progression through training in a time-independent fashion. What role SBA is to play in the context of competency-based training programmes[101] remains to be elucidated.

The development and implementation of a technical skills assessment into the European Board of Vascular Surgery certification process (Pandey et al, 2006) represented a key milestone in the history of simulation-based assessment in surgery[84]. The subsequent integration of the COSATS examination into the process of ABCRS certification, with validity evidence reported in line with a contemporary framework, is a more recent high-profile example of the use of simulation as an assessment modality at certification level[18]. De Montbrun et al. (2018) outline a systematic approach to the design, validation and implementation of a high-stakes assessment that can be adapted to further surgical specialties, institutions, and national curricula. Future studies should build on this model to explore the utility of SBA in other surgical specialties and training jurisdictions. How such an assessment fits into a multi-factorial certification process, and to what extent it complements other proxy assessments of skill such as trainer evaluations and operative experience logbooks, remains to be firmly established.

*Limitations*

The heterogeneity of study designs, populations, and outcome measures of included studies precluded meta-analysis. Simulation-based assessment literature pertaining to endoscopic or bedside procedures, as well as literature derived from performance assessments in non-surgical specialties, could provide further insight into the appropriate use of high-stakes technical performance assessments in surgery. These studies are beyond the scope of this review. Furthermore, assessments of aptitude or technical skill often form part of a wider programmatic approach to trainee

assessment; the validity, reliability, acceptability, cost effectiveness and educational impact of these assessments must therefore be analysed in context. Technical aptitude assessments and simulation-based assessments are resource intensive, somewhat limiting the ability of many training programmes and jurisdictions to implement an approach to trainee assessment outside of the operating theatre.

### 1.5 Conclusion

The development of contemporary validity evidence frameworks, competency-based education principles and higher-fidelity simulation has paved the way for the reliable, acceptable and valid use of simulation as a high-stakes assessment modality. This has culminated in the implementation of national simulation-based assessments of technical skill at certification level. Integrated approaches to simulation-based assessment, mapped to competency-based curricula, could see the successful development of a simulation-informed selection, progression, entrustment and certification process. Whether such a curriculum leads to more efficient training or improved patient safety remains to be explored by future studies. Given the observed gap between development or validation and implementation, it is likely that significant barriers to the deployment of simulation in high-stakes assessments remain; these should be explored and challenged.

## 1.6. Thesis Objectives and Research Questions:

Technical 'Performance assessments' are defined as assessments of operative or technical skill or competence that occur outside of the operating theatre[102]. In Irish surgical training, these include two distinct sets of assessments: 1) assessments of aptitude used to inform selection onto Core Surgical Training (the first two years of dedicated postgraduate surgical training) and progression from the first year of undifferentiated training to their second year of training in a specialty interest (ST1 to ST2). 2) Further simulation-based assessments of operative skill are conducted using simulation-based Objective Structured Clinical examinations (OSCES) at the end of each training year in Core Surgical Training. Scores awarded during these assessments contribute to an overall score used to rank trainees at the time of application to Higher Specialist Training (HST) in their chosen surgical specialty (Figure 1.6.1). Currently, there is no formal simulation-based assessment incorporated into Higher Specialist Training. At this level, assessment is primarily conducted in the workplace, as outlined by the curriculum of the Intercollegiate Surgical Curriculum Programme of the United Kingdom and Ireland (ISCP).

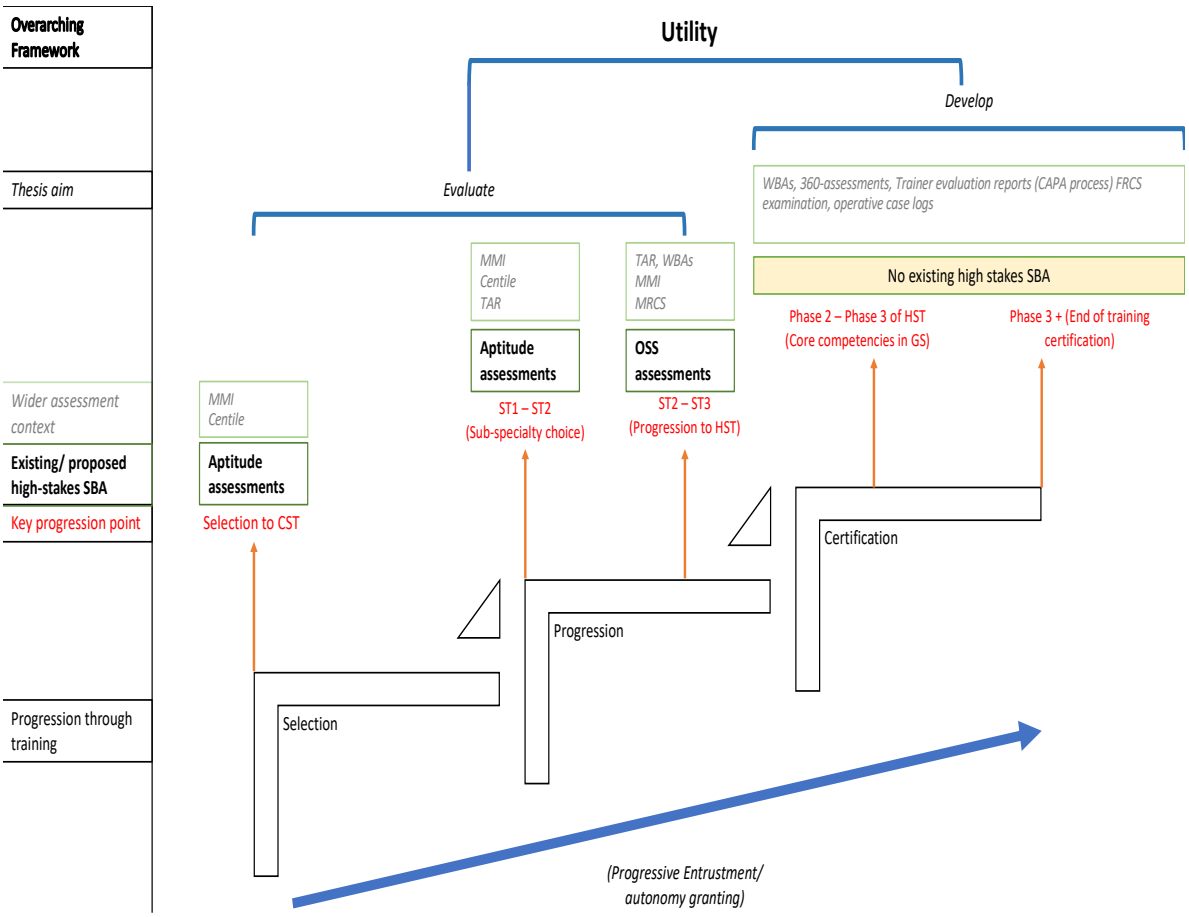



*Figure 1.6.1: Currently utilised and proposed technical performance assessments in Irish Surgical Training. SBA indicates Simulation-Based Assessment, MMI, Multiple Mini-Interview, CST, Core Surgical Training, TAR, Trainer Assessment Report, WBA, Workplace-Based Assessment, MRCS, Membership of the Royal College of Surgeons, ST, Surgical Training (year), OSS, Objective Surgical Skill, HST, Higher Specialist Training, CAPA, Competency Assessment and Performance Appraisal, FRCS, Fellowship of the Royal College of Surgeons, GS, General Surgery*

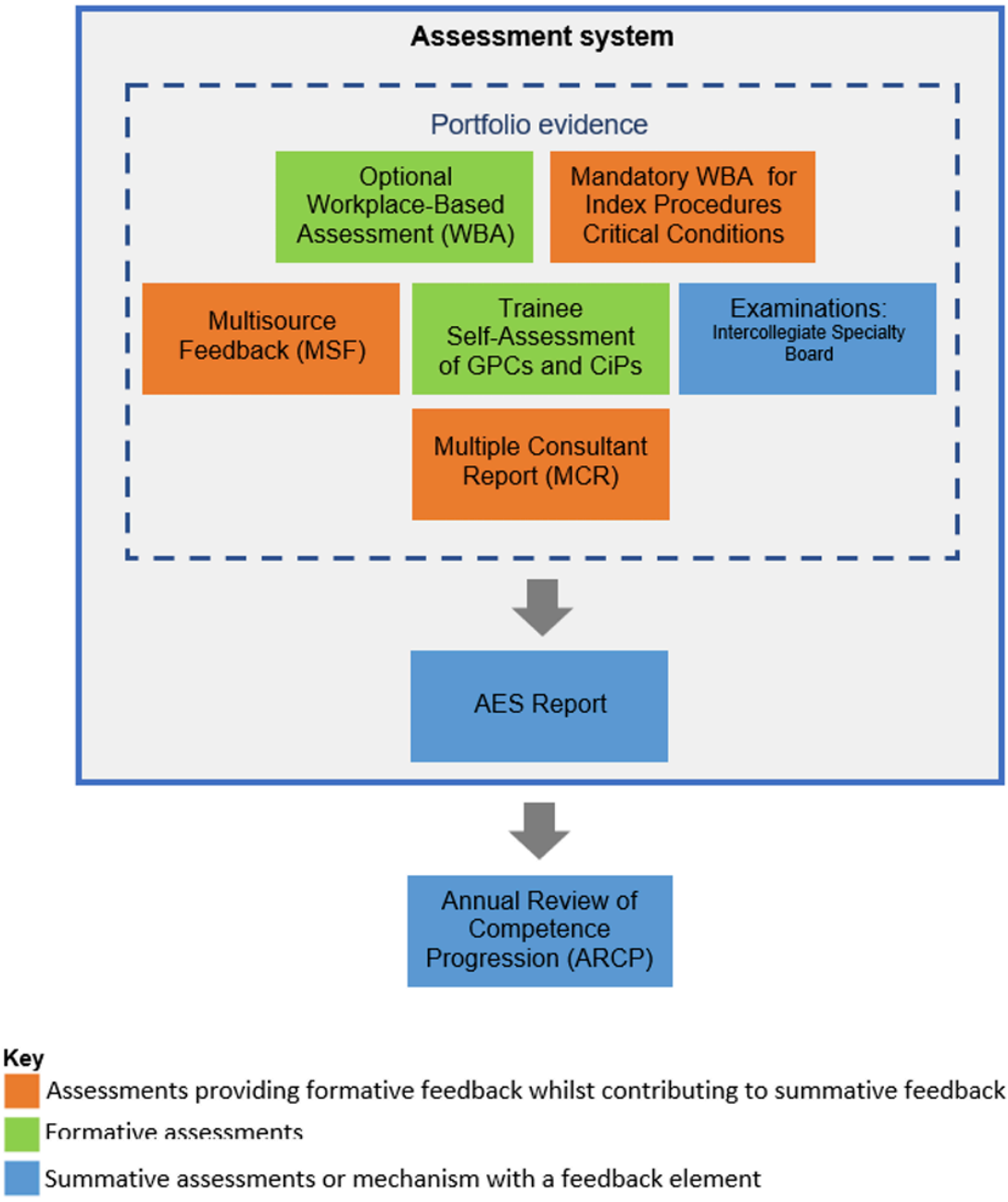



*Figure 1.6.2: Currently utilised assessment curriculum in General Surgery Higher Specialist Training. From the Intercollegiate Surgical Curriculum Programme, General Surgery Curriculum, 4 August 2021 (available at: https://www.iscp.ac.uk/iscp/curriculum-2021/, last accessed 26 June 2023). GPCs indicates Generic Professional Capabilities, CiPs, Capabilities in Practice, AES, Assigned Educational Supervisor*

The main body of this thesis concerns investigating the validity, reliability and acceptability of simulation as an assessment modality in three high-stakes scenarios in surgical training: at trainee selection, at progression from core surgical training to higher surgical training, and late training certification using general surgery as an index specialty. This study is conceptualised as an emergent, multiphase, mixed-method study (Figure 1.6.3). Concurrent phases, using qualitative and quantitative methodology, are used to evaluate the utility of currently utilised technical performance assessments in the Irish context. A further sequential stage seeks to generate utility evidence for the use of simulation as an assessment modality in higher specialist general surgery training. The domains of van der Vleuten's Utility framework are used to guide study design and data collation, i.e.: validity, acceptability, reliability, cost (or cost-effectiveness) and educational impact[22]. Within this framework, validity evidence is primarily collated according to Messick's five sources of validity evidence; content, internal consistency, relationships with other variables, response process and consequences[103]. Reliability findings are reported according to principles of classical test theory[104] and generalizability theory[105] where applicable, while education impact is measured according to Kirkpatrick's model as adapted for healthcare-related studies by Freeth et al[106].

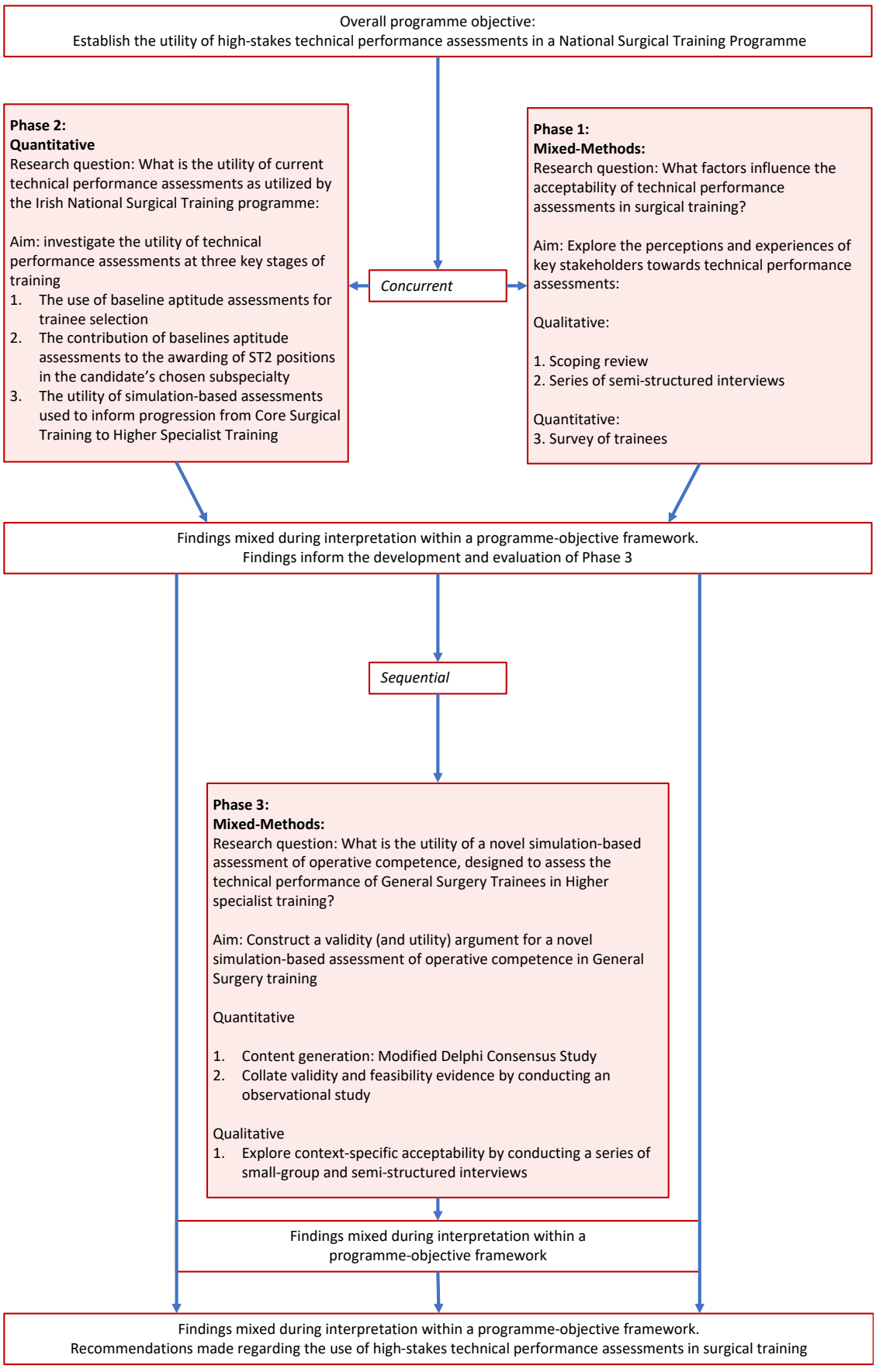



*Figure 1.6.3: Thesis conceptualised*

The corresponding objectives and principle research questions are summarised below:

**Objective 1:** To determine whether there is a need for simulation-based assessment in surgical training.

Research question: what is the added value of a simulation-based assessment curriculum in the context of contemporary, competency-based surgical training programmes?

**Objective 2:** To establish the validity evidence for the use of technical performance assessments, conducted outside of the operating room, in the selection of candidates to surgical training.

Research question: Are assessments of technical aptitude sufficiently validated and reliable for use in the recruitment of surgical trainees?

**Objective 3:** To explore the validity evidence for the use of simulation-based assessments of technical skill to inform decisions regarding the progression of surgical trainees from core surgical training to higher specialist training

Research question: Is the Objective Surgical Skills examination (OSS) sufficiently valid and reliable for use in the high-stakes assessment of surgical trainees, and how does it add utility to other, more traditionally utilised performance assessments such as operative experience logbooks and workplace-based assessments?

**Objective 4:** To develop and prospectively validate a novel simulation-based assessment of operative competence, designed to assess trainee in higher specialist training

Research question: Can a high-fidelity, multi-station technical skill assessment display sufficient validity and reliability metrics for use in the assessment of

competence in general surgery higher specialist training, and furthermore, what are the barriers to the implementation of such an assessment to inform end-of-training certification decisions?

**Objective 5:** To develop a proposed blueprint or framework regarding the optimum design, validation and implementation of an integrated, simulation-based assessment curriculum across the continuum of surgical training.

Research question: How should simulation be used as a high-stakes assessment modality across the continuum of surgical training, from selection to certification?

## 1.7. Data sources for this thesis

**Objective 1:**

For objective 1, the existing literature was explored in order to develop a moral argument for the use of simulation as a training and assessment tool in surgical training.

A series of studies were conducted to investigate the validity evidence for two key assessment modalities currently used by surgical training programmes across jurisdictions: workplace-based assessments and operative experience targets. A systematic (scoping) review was first conducted to map the current literature with respect to the assessment of technical skills Irish and United Kingdom surgical training. A further systematic review was conducted in order to specifically investigate the validity evidence supporting the use of operative experience targets or 'logbooks' in surgical training.

Subsequently, the acceptability of simulation as an assessment modality was explored in a three-stage process: first, a scoping review of the existing literature was conducted. Themes derived from thematic analysis of collated data were used to generate an interview guide, which was used to conduct a series of semi-structured interviews with key stakeholders in surgical training. Themes, sub-themes and codes derived from these semi-structured interviews were used to develop a survey for circulation across Ireland and the United kingdom, to explore the perspectives of surgical trainees regarding the use of simulation-based assessment in surgical training.

**Objective 2**

Prospectively collected data derived from applications to core surgical training was used to explore the predictive validity of baseline aptitude assessments as a selection tool, by exploring the associations between these assessment scores and scores achieved in future workplace-based and simulation-based assessments.

A further analysis of applicant data was conducted in order to explore the predictive validity of aptitude assessments with respect to likelihood of progression to higher specialist training, as part of a multi-modality selection framework including measures of undergraduate academic performance and performance at interview.

**Objective 3**

For objective 3, prospectively maintained trainee assessment data was retrospectively explored in order to generate validity and reliability evidence for workplace-based and simulation-based technical skill assessments in core surgical training.

A prospective validation study of the Objective Surgical Skills (OSS) assessment was further conducted.

**Objective 4:**

For objective 4, a prospective Delphi consensus study was first conducted in order to generate content validity for a novel simulation-based assessment of operative competence in higher surgical training. The proposed virtual reality for use in this novel assessment, the LapSim (Surgical Science, Sweden), was interrogated through a scoping review of the available literature regarding the use of this platform as an assessment and training tool. The validity, reliability and acceptability of the subsequently designed novel assessment was explored prospectively using a combination of quantitative and qualitative data points.

**Objective 5:**

A blueprint for the design, validation, and implementation of an integrated simulation-based assessment framework across the continuum of surgical training, from selection to end-of-training certification, was developed with respect to data derived from the above studies.

## Chapter 2: Determining the need for, and acceptability of, high-stakes technical performance assessments in surgical training

### Chapter overview:

This chapter consists of  a series of manuscripts exploring the need for, and acceptability of, simulation-based assessments in surgical training, with reference to the programmatic technical skill assessment framework of surgical training in Ireland and the United Kingdom. We begin by making a moral argument regarding the use of simulation-based training and assessment in surgical training. Two review articles are then presented. The first, 'Assessing Operative Skill in the Competency-based Education Era: Lessons From the UK and Ireland', presents the current assessment curriculum of the surgical training programme of Ireland and the United Kingdom, with specific reference to the utility of two primary methods of operative competence assessment; workplace-based assessments and operative experience targets for index procedures. The second, 'Defining operative experience targets in surgical training: A systematic review', further explores the utility of operative experience targets with reference to their use internationally. These reviews were conducted to determining the benefits and pitfalls of these operative assessment modalities, with a view to determining the relative merits of technical performance assessments outside of the operating theatre.

Two further published works are then presented, describing an iterative approach to determining the acceptability of technical performance assessments to key stakeholder representatives in surgical training. Finally, supporting unpublished data from a survey of Irish surgical trainees regarding the acceptability of such assessments is presented.

**Role of the thesis author:**

Mr Toale was the primary contributor to the study design and research question development for these studies, conducted the related literature searches, performed all data extraction and synthesis, drafted the initial manuscripts and produced the final versions. Furthermore, he conducted the series of semi-structured interviews, conducted the thematic analysis of transcribed data, developed the survey items for subsequent distribution, analysed the returned data, and drafted the draft and final manuscripts.

**References (published works contributing to this chapter):**

Toale C, Morris M, Kavanagh DO. Training to proficiency in surgery using simulation: is there a moral obligation? J Med Ethics. 2022 Jan 6:medethics-2021-107678. doi: 10.1136/medethics-2021-107678. Epub ahead of print. PMID: 34992083.

Toale C, Morris M, Kavanagh DO. Assessing Operative Skill in the Competency-based Education Era: Lessons From the UK and Ireland. Ann Surg. 2022 Apr 1;275(4):e615-e625. doi: 10.1097/SLA.0000000000005242. PMID: 35129531.

Toale C, O'Byrne A, Morris M, Kavanagh DO. Defining operative experience targets in surgical training: A systematic review. Surgery. 2022 Nov;172(5):1364-1372. doi: 10.1016/j.surg.2022.07.015. Epub 2022 Aug 26. PMID: 36038374.

Toale C, Morris M, Kavanagh D. Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review. Am J Surg. 2021 Oct;222(4):723-730. doi: 10.1016/j.amjsurg.2021.03.035. Epub 2021 Apr 6. PMID: 33849711.

Toale C, Morris M, Kavanagh DO. Perspectives on simulation-based assessment of operative skill in surgical training. Med Teach. 2022 Oct 26:1-8. doi: 10.1080/0142159X.2022.2134001. Epub ahead of print. PMID: 36288727.

***Additional material: further unpublished supporting works contributing to this thesis chapter:***

*Toale C, Morris M, Kavanagh DO. The Acceptability of Assessing Operative Competence Outside of the Operating Room: A Survey of Surgical Trainees (unpublished material).*

## 2.1: Training to Proficiency in Surgery using Simulation: Is there a moral obligation?

Reference: Toale C, Morris M, Kavanagh DO. Training to proficiency in surgery using simulation: is there a moral obligation? J Med Ethics. 2022 Jan 6:medethics-2021-107678. doi: 10.1136/medethics-2021-107678. Epub ahead of print. PMID: 34992083.

## 2.2. Assessing Operative Skill in the Competency-Based Education Era: Lessons from the UK and Ireland

Reference: Toale C, Morris M, Kavanagh DO. Assessing Operative Skill in the Competency-based Education Era: Lessons From the UK and Ireland. Ann Surg. 2022 Apr 1;275(4):e615-e625. doi: 10.1097/SLA.0000000000005242. PMID: 35129531.

## 2.3. Defining Operative Experience Targets in Surgical Training: A Systematic Review

Reference: Toale C, O'Byrne A, Morris M, Kavanagh DO. Defining operative experience targets in surgical training: A systematic review. Surgery. 2022 Nov;172(5):1364-1372. doi: 10.1016/j.surg.2022.07.015. Epub 2022 Aug 26. PMID: 36038374.

## 2.4. Perceptions and Experiences of Simulation-Based Assessment of Technical Skill in Surgery: A Scoping Review

Reference: Toale C, Morris M, Kavanagh D. Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review. Am J Surg. 2021 Oct;222(4):723-730. doi: 10.1016/j.amjsurg.2021.03.035. Epub 2021 Apr 6. PMID: 33849711.

## 2.5. Perspectives on simulation-based assessment of operative skill in surgical training

Reference: Toale C, Morris M, Kavanagh DO. Perspectives on simulation-based assessment of operative skill in surgical training. Med Teach. 2022 Oct 26:1-8. doi: 10.1080/0142159X.2022.2134001. Epub ahead of print. PMID: 36288727.

## *2.6. Additional Material: The Acceptability of Assessing Operative Competence Outside of the Operating Room: A Survey of Surgical Trainees*

**Abstract**

**Background**

Simulation is increasingly being explored as an assessment modality. There is little research exploring the perceptions of key stakeholders, including surgical trainees, toward simulation-based assessment (SBA). This survey-based study reports the perceptions and experiences of Irish surgical trainees toward SBA.

**Methods:**

An anonymous survey was developed using an iterative process of literature review and stake-holder interviews. This was distributed to all currently enrolled surgical trainees in Ireland, regardless of sub-specialty (N = 375). A mixed-method approach was adopted; quantitative data was collected via Likert-scale items, while qualitative data was collected via open-ended questions. Responses to open response comments underwent thematic analysis.

**Results:**

Fifty-four respondents completed the survey (response rate 14.4%). Simulation was perceived to be an appropriate method for assessing a trainee's knowledge of operative steps (mean 4.8 ± 1.16), and less useful in assessing higher-order skills such as communication and teamwork (3.34 ± 1.37) or decision making (2.94 ± 0.94). Lack of access to simulator technology and negative perceptions regarding available model fidelity were perceived barriers to the implementation of acceptable SBA.

**Conclusions**

Without significant investment in simulation-technology, SBA could be viewed by trainees as a barrier to training. SBA is most acceptably used to prepare and assess more junior trainees prior to real-world operating, as part of a train-to-proficiency

model. Barriers with respect to access, and the perceived lack of validity and fidelity of currently available simulator models, should be addressed before adoption of high-stakes SBA by training programmes.

### 2.6.1 Introduction

Simulation is an established tool in the field of surgical education and training, with the safe transfer of simulator-acquired technical skill to the operating theatre demonstrated in several studies[137 262 325]. The role of simulation in the assessment of technical skill proficiency is less firmly established. The Colorectal Objective Structured Assessment of Technical Skill (COSATS) and the General Surgery Objective Structured Assessment of Technical Skill (GOSATS) are the most notable examples of simulation-based assessment applied in a high stakes summative setting[14 15 18].

Prior studies suggest that surgical trainers view simulation as a valid method of assessment[271 272 276 278], while the perceptions of trainees regarding this concept are less well explored [271 280]. The broader application of simulation-based assessment in high-stakes areas such as trainee entrustment[51 69 272-274 276 277 279], certification[274 275] and trainee selection[272 275 326] is controversial. Consensus recommendations from the Association of Surgeons in Training of the United Kingdom in Ireland highlight concerns regarding the potential impact of simulation-based assessment on trainee clinical exposure[263]. The current model of in-theatre training and assessment is reported to be satisfactory from a patient safety perspective in light of evidence demonstrating no adverse outcomes related to trainee involvement in operative cases[128]. Further concerns exist regarding the perceived inability of current simulators to meet the complexity and fidelity requirements for high-stakes assessment[128].

The aim of this study is therefore to address this identified literature gap, by assessing trainee and trainer perceptions towards current simulation-based assessment in surgical training. Participant responses are intended to guide future iterations of simulation-based assessment curricula.

### 2.6.2 Methods

A survey tool was developed in line with the Association for Medical Education in Europe (AMEE) guide no. 87: Developing questionnaires for educational research (2014)[327]. A literature review was conducted in order to identify relevant prior research reporting the perceptions and experiences of trainees toward simulation-based assessment of technical or operative skill[328]. A series of semi-structured interviews were then recorded with a heterogenous group of stakeholder representatives using an interview tool developed using the previous literature review[329]. Themes, sub-themes and codes identified from this interview series were synthesized with findings from a literature review to create draft survey items. Expert content validation of the draft survey tool was carried out by a surgical trainer with extensive experience in simulation-based education, an expert surgical educator and researcher with experience in surgical simulation, and the Quality Enhancement Office of the Royal College of Surgeons in Ireland. Cognitive interviews were conducted with 10 surgical trainees, recruited using purposive sampling, using concurrent think-aloud interviewing and verbal probing techniques[330]. The final survey items were then compared to the themes and codes emerging from the initial semi-structured interview series, with an agreement of >80% considered to indicate content validity. The RCSI Quality Enhancement Office acted as a gatekeeper, distributing and collecting anonymous data. The full survey items are available (Appendix 2.6.1, supplemental data)

From each participant the following criteria were sought: surgical sub-specialty, level of training, and answers to questions provided, predominantly presented on a 6-point Likert scale to reduce central tendency, from 1 (strongly disagree) to 6 (strongly agree)[331]. Internal validity was ascertained using Cronbach's Alpha test of reliability, by analysis of questions pertaining to similar themes. A value of 0.6 or greater was considered to be an acceptable level of reliability. Quantitative analysis was performed using SPSS version 23 (IBM, California, USA). Responses to open response questions were analysed using reflexive thematic analysis[332], according to an inductive, constructivist framework. A brief reflexivity statement follows: CT (first

author) completed the Royal College of Surgeons in Ireland (RCSI) Core Surgical Training (CST) programme and is currently a higher specialist trainee (HST) in General Surgery. Annual simulation-based assessments of technical and non-technical skill are part of the assessment curriculum at CST level; these assessments inform progression through to HST in a trainee's chosen sub-specialty. CT has further undertaken a higher research degree in technical skill assessment, to which this work contributes. MM is a senior lecturer within RCSI, and has applied and research expertise in surgical education. DOK is currently Head of Surgical Research within RCSI, is a trainer of higher specialist and core surgical trainees, and is a general and colorectal surgeon. The authors are therefore familiar with the use of simulation as an assessment modality, and in particular with the simulation-based training and assessment programme administered by the Royal College of Surgeons in Ireland at the National Surgical Training Centre.

### 2.6.3 Results:

Fifty-four trainees completed the survey (response rate 14.4%). Mean responses to Likert scale items are outlined in Table 2.6.1.

*Table 2.6.1: Mean responses to Likert items (1-6), where 1 = strongly disagree and 6 = strongly agree.*

| Item | Mean | SD |
|---|---|---|
| I am confident that when I obtain my certificate of completion of training (CCT), I will be competent in performing the expected operative procedures in my specialty without assistance, including when managing complications | 4.51 | 1.25 |
| Current methods of operative competency assessment (procedure-based assessments, multiple consultant reports, multi-source feedback and indicative number targets) are satisfactory as summative assessments of operative competence | 3.09 | 1.04 |
| Using simulation to assess operative competence is more objective than assessments performed in the operating theatre | 3.00 | 1.19 |
| Using simulation to assess operative competence is easier to standardise than assessments performed in the operating theatre | 4.03 | 1.27 |
| Simulation-based assessment is a more constructive way of receiving feedback or giving feedback to trainees than in-theatre assessments | 2.91 | 1.12 |
| Using simulation-based assessment to ensure trainees meet the minimum level of operative competency expected for their stage of training would allow in-theatre training time to be used more efficiently | 3.97 | 1.25 |
| A trainee's performance in a simulated operative procedure is reflective of their performance in the operating theatre | 3.00 | 0.77 |
| Simulation is an accurate method of assessing whether a trainee knows the steps of a procedure | 4.80 | 1.16 |
| Simulation is an accurate method of assessing a trainee's operative technique or skill | 3.37 | 0.94 |
| Simulation is an accurate method of assessing a trainee's non-technical skills such as communication and teamwork | 3.34 | 1.37 |
| Simulation is an accurate method of assessing a trainee's ability to perform an operative procedure under pressure | 2.49 | 0.85 |
| Simulation is an accurate method of assessing a trainee's intra-operative decision-making skills | 2.94 | 0.94 |
| Simulation-based assessment could be used to determine whether trainees meet the minimum operative competency level expected for their stage of training | 3.57 | 1.17 |
| Simulation-based assessment of operative competence can identify under-performing trainees | 3.49 | 1.19 |
| Simulation-based assessments can distinguish between trainees who meet operative competency expectations and trainees who exceed expectations | 2.89 | 0.90 |
| Using simulation-based assessment to ensure trainees meet the minimum operative competency level expected for their stage of training would have a positive impact on patient safety | 3.76 | 1.26 |
| Simulation-based assessment can accurately replicate the real-world operating theatre environment | 2.74 | 0.78 |
| Performance in simulation-based assessment can provide objective evidence of a trainee's progression in operative competence over time | 3.83 | 1.22 |
| A human assessor is a better determinant of a trainee's ability than simulator-measured metrics (i.e. metrics recorded by a computer) | 4.20 | 1.16 |
| Simulator-measured metrics, recorded and assessed by a computer, are accurate methods of assessing operative competence | 2.91 | 0.79 |
| A trainee's 'logged hours' spent using a simulator is reflective of a trainee's operative competence | 2.49 | 0.82 |
| If simulation were to be used to assess a trainee's operative competence, the models would need to have a very realistic appearance | 5.45 | 0.79 |
| If simulation were to be used to assess a trainee's operative competence, the models would have to reflect real-world case variety | 5.42 | 0.79 |

| | | |
|---|---|---|
| If simulation were to be used to assess a trainee's operative competence, it would be important for the simulation to reflect the real-world theatre environment | 5.24 | 1.06 |
| If simulation were to be used to assess a trainee's operative competence, it would be important for the simulated models to handle like human tissue | 5.69 | 0.47 |
| The more senior the trainee being assessed, the more realistic the simulator model should be | 4.27 | 1.82 |
| Low-fidelity synthetic bench models (e.g. laparoscopic box trainers or bench-model suturing kits) are suitable for use in assessments of a trainee's operative competence | 2.72 | 0.76 |
| High-fidelity synthetic bench models (synthetic replicas of anatomical models for surgical procedures) are suitable for use in assessments of a trainee's operative competence | 3.52 | 1.12 |
| Animal models are suitable for use in assessments of a trainee's operative competence | 3.58 | 1.22 |
| Virtual reality or computer-generated platforms (e.g. robotic/ laparoscopic/ endovascular platforms) are suitable for use in assessments of a trainee's operative competence | 3.48 | 1.06 |
| Cadaveric models are suitable for use in assessments of a trainee's operative competence | 4.15 | 1.30 |
| Simulation-based assessment is suitable for assessing a trainee's competence in open surgery | 3.00 | 0.00 |
| Simulation-based assessment is suitable for assessing a trainee's competence in Endoscopic surgery | 3.43 | 1.13 |
| Simulation-based assessment is suitable for assessing a trainee's competence in Robotic surgery | 3.71 | 1.25 |
| If simulation were to be used to assess operative competence, it would suffice to simulate key steps of the procedure, and not necessarily the whole procedure | 3.77 | 1.20 |
| Simulation should be used to assess competence on a series of core, commonly encountered procedures | 4.16 | 1.19 |
| Simulation should be used to assess competence in rarely encountered, high-risk procedures | 3.87 | 1.61 |
| Technical ability is an important consideration when selecting trainees for higher specialist training | 4.23 | 1.33 |
| Using simulation-based assessment of operative competence to inform selection decisions for higher specialist training is fair | 3.19 | 1.30 |
| A trainee's performance in simulation-based assessment should be used to decide how much autonomy trainees are granted in the operating theatre | 2.71 | 0.82 |
| Trainees should have to demonstrate competence on a simulated procedure before they can perform that procedure on a patient under supervision as the primary operator | 2.77 | 1.02 |
| Using simulation-based assessment to decide how much autonomy trainees are granted in the operating room would result in less operative exposure for trainees | 4.42 | 1.20 |
| Using simulation-based assessment to decide how much autonomy trainees are granted in the operating theatre would positively incentivize trainees to improve their operative performance | 3.26 | 1.21 |
| Using simulation-based assessment to decide how much autonomy trainees are granted in the operating room would positively impact upon patient safety | 2.97 | 0.98 |
| Using simulation-based assessment to decide how much autonomy trainees are granted in the operating room would improve trainee confidence | 3.32 | 1.19 |
| Simulation-based assessment should replace a trainer's expert opinion regarding the appropriate amount of autonomy to give to a trainee | 2.00 | 0.97 |
| It is fair to use scores in simulation-based assessment, in conjunction with other assessments, to inform decisions on a trainee's progression to the next training year | 3.13 | 1.06 |
| Performance in simulation-based assessment of operative competence alone is insufficient when making decisions regarding a trainee's progression to the next training year | 4.77 | 1.54 |
| It is more difficult to assess the operative competence of more senior trainees using simulation | 4.39 | 1.33 |

| | | |
|---|---|---|
| A trainee's operative competence should be formally assessed at the end of training | 4.62 | 1.27 |
| Simulation-based assessment could be used to inform decisions regarding a trainee's operative competence at the end of training, in conjunction with other assessments | 3.39 | 1.42 |
| It is more ethical, from a patient's perspective, to use simulation-based assessment than in-theatre operative cases to assess trainees at the end of training | 3.00 | 1.00 |
| Using simulation-based assessment to assess a trainee's operative competence at the end of training is fair | 3.14 | 1.11 |
| Using simulation-based assessment alone is sufficient when making decisions regarding a trainee's operative competence at the end of training | 1.68 | 0.86 |
| Simulators that I have experienced are suitable for use in simulation-based assessments informing decisions regarding selecting trainees for higher specialist training | 2.43 | 0.99 |
| Simulators that I have experienced, are suitable for use in informing decisions regarding how much autonomy trainees are granted in the operating theatre | 2.25 | 1.04 |
| Simulators that I have experienced are suitable for use in informing decisions regarding the progression of trainees through higher specialist training | 2.32 | 1.02 |
| Simulators that I have experienced are suitable for use in assessing a trainee's operative competence at the end of training | 1.96 | 0.96 |

SD indicates Standard Deviation

Trainees were confident that when they complete training, they will be competent performing the expected operative procedures in their specialty without assistance, including when managing complications (mean 4.51 ± 1.25), though were neutral as to whether current methods of operative competency assessments were satisfactory (3.09 ± 1.04). While respondents reported that simulation is an accurate method of assessing whether a trainee knows the steps of a procedure (4.80 ± 1.16), but were more neutral as to whether simulation was an accurate method of assessing a trainee's operative technique or skill (3.37 ± 0.94) or non-technical skills such as communication or teamwork (3.34 ± 1.37). Trainees were less likely to agree that simulation is an accurate method of assessing a trainee's ability to perform an operation under pressure (2.49 ± 0.85), or their intra-operative decision-making skills (2.94 ± 0.94).

Trainees value model fidelity highly. Trainees responded that if simulation were to be used to assess a trainee's operative competence, the models would need to have a very realistic appearance (5.45 ± 0.79), and that models would have to reflect real-world case variety (5.42 ± 0.79). Furthermore, trainees reported that if simulation were to be used to assess a trainee's operative competence, it would be important for the simulation to reflect the real-world theatre environment (5.24 ± 1.06) and that it would be important for the simulated models to handle like human tissue (5.69 ± 0.47). Trainees were sceptical regarding the use of SBA in autonomy granting decisions, disagreeing with the statement "Simulation-based assessment should replace a trainer's expert opinion regarding the appropriate amount of autonomy to give to a trainee" (2.00 ± 0.97). While trainees agree that a trainee's operative competence should be formally assessed at the end of training (4.62 ± 1.27), they were sceptical regarding the use of simulation for this purpose. Trainees disagreed with the statement: Simulators that I have experienced are suitable for use in assessing a trainee's operative competence at the end of training (1.96 ± 0.96). Responses to further Likert items are outlined in Table 2.6.1.

Themes, sub-themes and exemplary quotes derived from thematic analysis from open response comments are outlined in Table 2.6.2.

*Table 2.6.2: Themes, sub-themes and exemplary quotes generated from thematic analysis of open response comments:*

| Theme | Sub-theme | Exemplary Quote |
|---|---|---|
| Access as a barrier: Without significant investment in simulator technology, SBA could be seen as a barrier to training | SBA is limited significantly by access to available resources | "I have concerns about facilities to perform these and there being significant backlog preventing trainees from progressing |
| | SBA could negatively impact on real-world assessment and training opportunities | "May reduce real world teaching and assessment." |
| Perceived deficiencies in reliability: While SBA may be more standardisable than in-theatre assessments, the use of an unfamiliar assessor and a single time point impacts negatively on its perceived reliability | One-off SBA is less reliable than repeated in-theatre assessment | "should not replace regular ongoing assessment and opinion of the consultant trainers who will have worked with the trainee in a wide variety of settings over the course of the rotation." |
| | SBA is more reliable or standardisable than in-theatre assessments | "Can standardize procedures. Although surgical procedures are rarely standard." |
| Perceived deficiencies in Validity: The acceptability of SBA is limited by negative perceptions regarding its validity | The fidelity of currently available simulators negatively impacts trainee's perceptions of their validity | "I have not encountered tissue models that I think would be suitable to make high stake (i.e. progression/sign off) decisions on, and that includes cadaveric simulation". |
| | Trainees perceived a significant lack of construct validity related to SBA | "usually not reflective of real-life operative experience and therefore more appropriate as a training tool rather than for assessment […] should not replace regular ongoing assessment and opinion of the consultant trainers who will have worked with the trainee in a wide variety of |

| | | |
|---|---|---|
| | | settings over the course of the rotation. |
| Train-to-proficiency: SBA would be acceptable for use prior to real world operating as part of a training to proficiency model | SBA can be useful for assessing performance in rarely encountered procedures | "Provides exposure to less commonly seen scenarios allowing for greater confidence if one finds themselves in that situation in vivo" |
| | SBA is more useful for formative, rather than summative assessment | "Allows for highlighting of areas where knowledge or skills are lacking." |
| | SBA is more useful in assessing trainees at an early stage of training | "Simulation likely helpful in early stages of training ?Less helpful as training progresses" |
| | SBA is safe for patients | "Low risk environment. Builds confidence. Allows for development of skills and improves patient safety". |
| | SBA is useful in assessing whether trainees know the steps of a given procedure | "You can get insight into whether or not the trainee knows the steps of the procedure". |
| | Simulation is more useful as a teaching modality than an assessment modality | " I think simulation is a very useful training tool: it should be formative not summative and I fundamentally disagree with using it in high stakes assessments". |
| | Simulation-based training to proficiency improves trainee confidence prior to real-world performance of procedures | "Simulation was very helpful when I was faced with doing a chest drain at 2am in the morning. I was able to recall the steps and perform the procedure safely." |
| | simulator-led proficiency-based progression could allow in-theatre training to be used more efficiently | "more likely for trainee to actually do all of procedure rather than trainer taking over in interest of time". |

SBA indicates Simulation-Based Assessment

Respondents were asked to respond to two open-response items: “What concerns do you have regarding the use of simulation-based assessment of operative competence in surgery?” and “What are the positive aspects of simulation-based assessment of operative competence in surgery?”. Sixty-nine unique responses were analysed. It emerged that limited access to simulators was perceived as a significant barrier to its use as an assessment modality. Trainees had ‘concerns about facilities to perform [SBA] and there being significant backlog preventing trainees from progressing’. This concern regarding progression was further reflected in responses highlighting the potential negative impact of SBA on real-world training opportunities, in particular that it “may reduce real world teaching and assessment”. While it was acknowledged by some respondents that SBA is easier to standardise when compared to workplace-based assessment, its use as a summative modality was hampered by the infrequency of SBA and the benefits of workplace-based observation in providing a longitudinal assessment of a trainee’s ability: “[SBA] should not replace regular ongoing assessment and [the] opinion of the consultant trainers who will have worked with the trainee in a wide variety of settings over the course of the rotation.” Of further concern to respondents was the perceived lack of fidelity of currently available simulators. One respondent commented - “I don’t think any simulation model currently adequately reflects operative skill sufficiently to make high stakes judgements on trainees based on them”. This was coupled with a belief that SBA was “usually  not reflective of real-life operative experience”.

In spite of these negative perceptions and experiences of simulation-based assessment, trainees offer examples of instances where SBA could add value to surgical training (Figure 2.6.1).

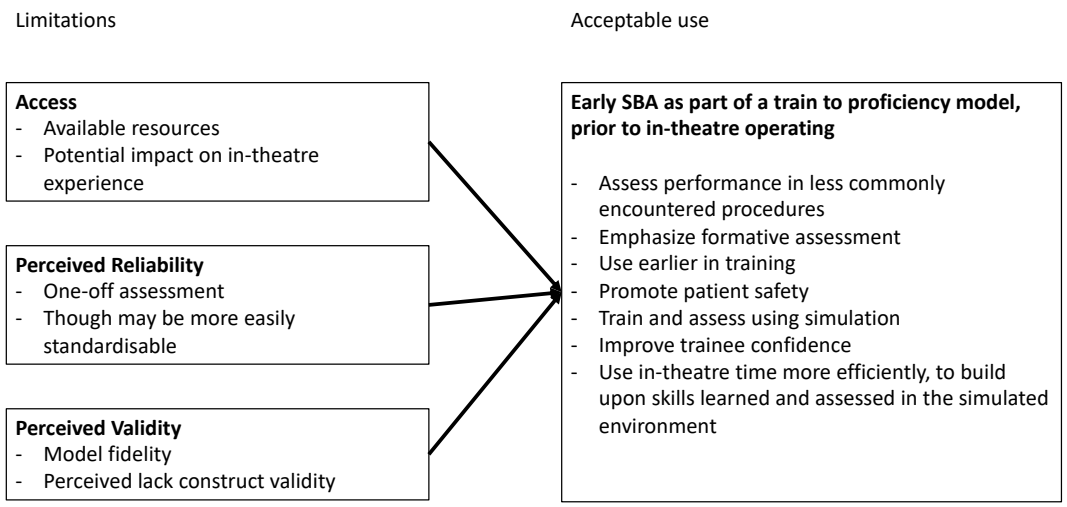



*Figure 2.6.1: Limitations of SBA, and methods for promoting its acceptable use in surgical training, generated from thematic analysis of open-response survey items.*

On thematic analysis, it emerged that SBA may be most useful in early surgical training, prior to in-person operating as part of a train-to-proficiency model, with a view to improving trainee confidence, promoting patient safety, and allowing in-theatre training time to be used more efficiently. SBA was reported to provide "exposure to less commonly seen scenarios", and "allows for highlighting of areas where knowledge or skills are lacking". Its acceptable use in early surgical training appears coupled to the idea that SBA is most useful in gaining "insight into whether or not the trainee knows the steps of the procedure".

### 2.6.4 Discussion:

Findings from this survey of Irish trainees suggest that simulation-based assessments are perceived to be more readily standardisable than workplace-based assessment. While trainees are confident regarding the ability of SBA to assess procedural knowledge, they are less so regarding its ability to determine competence in higher order skills such as communication and decision making. This may be associated with a perceived inability of SBA to identify high-performing trainees. Model fidelity is highly valued by trainees; acceptable simulations for assessment

purposes are likely those which closely approximate the complexity of real-world cases and the operating environment. Fidelity is perceived as being particularly important when assessing more senior trainees. Cadaveric models are perceived to be the most suitable for use in assessment. Trainees have concerns that increasing reliance on SBA could lead to reductions in real-world operative experience, and do not support its sole use in autonomy-granting decisions. Performance in SBA is deemed to be insufficient in deciding on progression through training when used in isolation. While trainees believe that operative skill should be formally assessed at the end of training, they do not believe that performance in SBA is sufficient when determining operative competence at this late training stage. Irish trainees do not believe that simulations they have encountered as part of their training are appropriate for use in making high-stakes training decisions such as trainee selection, autonomy granting, progression through training or end-of-training certification.

The perceptions of key stakeholders regarding the value of simulation-based assessment of operative competence varies in the published literature. A number of survey-based studies have identified positive perceptions of programme directors across surgical disciplines toward simulation as an assessment modality[270 271 276 278]. Residents in further studies also report that simulation is a fair and effective method of assessment[271], acknowledging its potential role in improving patient safety[280]. Residents and programme directors alike are less certain of its role in informing autonomy-granting decisions[333], end-of training certification[92], or trainee selection[41]. This current study further explores the perceptions and experiences of trainees towards SBA, suggesting targets for improvement in assessment in order to improve acceptability. A train-to-proficiency model could be employed, by exposing trainees to simulation early in training. Trainees could be trained and assessed to meet a pre-defined standard prior to undertaking real-world operating. The formative aspect of SBA could be emphasised, with assessment conducted to promote patient safety, improved trainee confidence, and ensure that subsequent in-theatre training time was used more efficiently. Simulation-led training and assessment could provide an

opportunity for supervised, proficiency based deliberate practice in line with principles of spaced learning[334].

This study is limited by its sample size. Responses are from surgical trainees only and therefore may not represent the views of other key stakeholders such as surgical trainers and patients. This survey, however, provides detailed data regarding the factors which influence the acceptability of simulation as an assessment modality for surgical trainees, highlighting significant barriers that will need to be addressed prior to the adoption of such assessment modalities in high-stakes training decisions such as those of trainee selection, progression, entrustment and certification.

## Chapter 3: Validity evidence for objective technical aptitude assessment at the time of surgical trainee selection

This chapter outlines the validity evidence for the objective assessments of technical aptitude used by the Royal College of Surgeons in Ireland's (RCSI) multi-factorial selection process.

Role of the thesis author:

Mr Toale was the primary contributor to the study design and research question development for these studies, conducted the statistical analyses, drafted the initial manuscripts and produced the final versions.

References: (published works contributing to this chapter):

Toale C, Morris M, Ryan DM, Boland F, Doherty EM, Traynor OJ, Kavanagh DO. Baseline Assessments of Psychomotor Aptitude are Associated with the Future Operative Performance of Surgical Trainees A Prospective Observational Cohort Study. Ann Surg. 2022 Jul 15. doi: 10.1097/SLA.0000000000005565. Epub ahead of print. PMID: 35837886.

Additional Material (unpublished work contributing to this chapter):

Toale C, Morris M, Gross S, O'Keeffe D, Ryan DM, Boland F, Doherty E. Traynor OJ, Kavanagh DO. Performance at National Selection and Future Performance in Surgical Training (under review, JAMA Surgery)

## 3.1 Baseline assessments of psychomotor aptitude are associated with the future operative performance of surgical trainees

**Reference:** Toale C, Morris M, Ryan DM, Boland F, Doherty EM, Traynor OJ, Kavanagh DO. Baseline Assessments of Psychomotor Aptitude are Associated with the Future Operative Performance of Surgical Trainees A Prospective Observational Cohort Study. Ann Surg. 2022 Jul 15. doi: 10.1097/SLA.0000000000005565. Epub ahead of print. PMID: 35837886.

## 3.2 Additional Material: Performance at National Selection and Future Performance in Surgical Training

Reference: Toale C, Morris M, Gross S, et al. Performance in Irish Selection and Future Performance in Surgical Training. JAMA Surg. Published online March 06, 2024. doi:10.1001/jamasurg.2024.0034

## Chapter 4: Exploring the utility of an implemented simulation-based assessment designed to inform decisions on progression from Core Surgical Training to Higher Specialist training

This chapter consists of  a series of manuscripts exploring the ways in which simulation-based training can integrate with, and complement, a workplace-based assessment curriculum. We first explore two identified pitfalls related to workplace-based assessment of operative competence in the context of Irish Core Surgical Training: 1) Response process validity (i.e. the way in which workplace-based assessments are engaged with in a real-world context) and 2) the reliability of workplace—based assessment, particularly related to their use in determining a 'global' impression of a trainee's operative skill or competence level.

In further supporting (unpublished) work presented herein, we explore the validity of an established assessment in Irish Core Surgical Training – the Operative Surgical Skill (OSS) examination, which contributes to decisions regarding progression from Core Surgical Training to Higher Specialist Training in the trainee's chosen surgical specialty.

Note: The validity and reliability of RCSI's approach to programmatic assessment in Core Surgical Training has previously been explored by Ms. Dara O'Keeffe, Simulation Lead at the Department of Surgical Affairs at RCSI. Ms O'Keeffe's MHPE thesis is available at

https://indigo.uic.edu/articles/Evaluating_the_Validity_of_a_National_Multi-Assessment_System_in_Postgraduate_Surgical_Training/12481124/files/23145071.pdf

Role of the thesis author:
Mr Toale was the primary contributor to the study design and research question development for these studies, performed all data extraction and synthesis, drafted the initial manuscripts and produced the final versions.

References: (published works contributing to this chapter):

Toale C, Nally DM, Ryan DM, Morris M, Kavanagh DO. How do Trainers and Trainees use Formative Workplace-based Assessments of Operative Competence to Deliver and Receive Structured Feedback? J Surg Educ. 2022 Mar-Apr;79(2):485-491. doi: 10.1016/j.jsurg.2021.08.023. Epub 2021 Sep 28. PMID: 34593328.

Toale C, Morris M, O'Keeffe D, Boland F, Ryan DM, Nally DM, Kavanagh DO. Assessing operative competence in core surgical training: A reliability analysis. Am J Surg. 2023 Jun 30:S0002-9610(23)00273-8. doi: 10.1016/j.amjsurg.2023.06.020. Epub ahead of print. PMID: 37481408.

Additional Material (unpublished work contributing to this chapter):

Toale C, Morris M, Kavanagh DO, O'Keeffe D. Prospective Validation of the Operative Surgical Skill (OSS) Examination: A Simulation-Based Assessment of Operative Competence for Core Surgical Trainees

Toale C, Morris M, Kavanagh DO, O'Keeffe D. Comparing the Reliability of Simulation-Based and Workplace-Based Assessments of Operative Competence in Core Surgical Training

## 4.1 Factors negatively impacting upon the reliability of workplace-based assessments of operative competence

Reference: Toale C, Nally DM, Ryan DM, Morris M, Kavanagh DO. How do Trainers and Trainees use Formative Workplace-based Assessments of Operative Competence to Deliver and Receive Structured Feedback? Journal of Surgical Education. 2022 Mar-Apr;79(2):485-491. DOI: 10.1016/j.jsurg.2021.08.023. PMID: 34593328.

## 4.2 Exploring the Number of Observations Required to Reliably Assess Operative Competence in Core Surgical Training

Reference: Toale C, Morris M, O'Keeffe D, Boland F, Ryan DM, Nally DM, Kavanagh DO. Assessing operative competence in core surgical training: A reliability analysis. Am J Surg. 2023 Nov;226(5):588-595. doi: 10.1016/j.amjsurg.2023.06.020. Epub 2023 Jun 30. PMID: 37481408.

## 4.3 Additional Material: Prospective Validation of the Operative Surgical Skill (OSS) Examination: A Simulation-Based Assessment of Operative Competence for Core Surgical Trainees

**Abstract**

**Introduction**

Simulation represents a standardisable approach to operative performance assessment. This study sought to prospectively collate validity evidence for the use of the Objective Surgical Skill (OSS) assessment in Core Surgical Training (CST), in the context of the national surgical training program of the Royal College of Surgeons in Ireland (RCSI).

**Methods**

Results of an annual sitting of the OSS assessment were analysed and collated according to Messick's five sources of validity evidence. Analysis was conducted using SPSS software version 23 (IBM, California, USA).

**Results**

A total of 93 candidates undertook a four station operative skills assessment. The inter-rater reliability was acceptable (0.71) for the total score but poor for the global competence rater (0.49). Performance across 20 stations would need to be observed for the OSS examination to be reliably used as a summative measure of operative skill for high-stakes decisions. The station pass rate varied from 79.57% to 90.32% across stations.

**Conclusion**

Simulation-based assessment is feasible and can inform decisions regarding trainee progression as part of a programmatic approach to trainee assessment.

### 4.3.1 Introduction

The traditional Halstedian of apprenticeship training in surgery has undergone significant revolution in recent decades[1]. Evolving patient expectations regarding the role of surgical trainees in their care[2], an increased emphasis on theatre efficiency[3], and well explored concerns regarding perceived operative competence and confidence of graduating trainees[455], has led to a re-evaluation of the training paradigm. 'Competency-based' approaches to outcome-driven training and assessment are well established across training jurisdictions[374 456], leading to the development of nominally time-independent postgraduate surgical programs[9]. The surgical specialties pose a unique challenge in this regard, due to the requirement for robust and reliable methods of teaching and assessing competence in operative skill[10].

The role of simulation in surgery has evolved in tandem with competency-based education[12]. Simulation is commonplace in surgical training curricula [13], and is increasingly being deployed as a method of assessing a trainee's competence. Studies have demonstrated the safe transfer of simulator-acquired operative skill to the operating theatre [16]. Whether simulation-training ultimately translates to improved patient safety and surgical outcome measures has not yet been firmly established [17]. Simulation has most notably been used in high-stake assessments through the implementation of the Colorectal Objective Structured Assessment of Technical Skill by the American Board of Colon and Rectal Surgery[18]. The Royal College of Surgeons in Ireland uses a multi-modal assessment framework to decide on trainee progression to higher specialist training in a surgical sub-specialty. This framework consists of workplace-based assessments, operative experience logbooks, end-of-rotation trainer evaluation reports, and a multi-station interview. Bothe technical and non-technical skills are also assessed using multi-station OSCE-style assessments. One of these assessments, the 'OSS' or Operative Surgical Skill' assessments, is a multi-station technical skill examination where core surgical trainees are assessed performing procedures or part-tasks using animal or synthetic models.

We hypothesise that the OSS is s a valid method of operative competence assessment for trainees undertaking core surgical training. This study seeks to prospectively generate validity evidence for a simulation-based assessment of operative competence used by the Royal College of Surgeons in Ireland to assess the operative competence of Core Surgical Trainees (CSTs). This assessment is designed for use at key stage of training; the transition from undifferentiated surgical training to further sub-specialty training. This study seeks to describe and collect validity evidence for the use of this assessment tool in accordance with Messick's validity framework. Messick describes five sources of validity evidence: Messick's framework outlines five sources of validity evidence[171]. 'Content validity' describes relationship between the content of an assessment and the construct it is intended to measure[457]. The 'Internal Structure' validity includes evidence on the assessment's internal consistency, as well as inter-rater reliability[458]. 'Relationships with other variables' are analysed via correlations with other assessments measuring operative competence, and traditionally using novice-expert comparisons[457]. 'Response process' evidence often consists of analyses of assessor or resident thoughts or actions during the assessment, assessment security or quality control, and evidence on rater training[457]. Evidence of the 'consequences' of an assessment refers to both beneficial and harmful effects on trainee performance, clinical outcomes and other trainee effects such as unintended stress or anxiety[459].

### 4.3.2 Methods

#### *4.3.2.1 Study design, setting and participants*

This study is reported according to the STrengthening the Reporting of OBservational studies in Epidaemiology (STROBE) guidelines[349]. The OSS is a 4-station assessment which is undertaken at two time-points during CST; at the end of each training year (CT1 and CT2). This assessment is also used to assess candidates of the Royal College of Surgeons in Ireland Master's in Surgical Science and Practice (MSSP) programme, which is undertaken by enrolled candidates prior

to undertaking dedicated post-graduate surgical training in the hospital context. These students therefore have limited real-life operative experience, though they undergo a simulation-based training curriculum in operative skill within RCSI. All candidates rotate through each OSS station which is assessed by a single examiner using pre-defined procedure- or task- specific checklist, and the well-established OSATS tool[460]. Both of these tools are equally weighted and combined to produce a final score for each station. Procedures or tasks are derived from the simulation-based training curriculum, and represent tasks and models which the candidates have been previously familiarized with, and which are deemed suitable for trainees at this level. In this iteration, stations were: Example checklists are provided (Appendix 4.3.1, supplemental data), A global, binary 'competent/ not competent' rater is also provided. This iteration of the OSS assessment recruited CSTs in their first year of training, and candidate of RCSI's MSSP programme. The assessment took place in the National Surgical Training Centre of the Royal College of Surgeons in Ireland. All candidates partaking in this assessment in April 2022 were eligible for inclusion. Examiners and candidates were familiarized with the assessment tools and models prior to the date of assessment. Synthetic and animal models are those encountered by Core Surgical Trainees and MSSP students during their respective simulation-based teaching curricula. Ethical approval was granted by the University of Medicine and Health Sciences at the Royal College of Surgeons in Ireland.

#### *4.2.2.2 Study outcome measures*

*Internal structure validity:* Internal consistency reliability will be assessed by comparison of scores achieved in the OSS checklists with the global competency score. This will be performed using Spearman's rho and Pearson correlation coefficients. Cronbach's alpha will be used to assess for inter-item agreement between checklist items, providing further internal consistency validity for the OSS Assessment tool. A random selection of ___ bowel anastomosis stations were assessed by a second independent rater. Inter-rater reliability will be assessed using intra-class coefficients, and further explored using Generalizability theory (see detailed description under 'statistical analysis'). Generalizability theory will be used to

explore the reliability of the OSS assessments. This first involves calculating the contribution of identified variables to overall score variance; this will be performed using the VARCOMP procedure in SPSS. Identified contributing variables are as follows: trainee variance, inter-station variance, cohort variance, error. The 'true' or 'intended' variance is then calculated for each assessment, and refers to the proportion of variance attributable to the trainee (i.e. due to varying ability of different trainees). This is calculated as *(Trainee variance/ Trainee variance + all other sources of variance)* and is called the 'reliability co-efficient' or '*G*'. *G* ranges from 0 – 1, where 1 represents perfect reliability[435]. Simple mathematical modelling can then be performed in order to model *G* with increasing numbers of hypothetical stations. This is called a Decision study or 'D-study'[435]. A fully crossed study design will be used. Trainee and stations will be considered as a random sample from an infinite universe.

*Relationships with other variables:* Scores achieved will be compared to overall procedural numbers, as recorded in each trainee's eLogbook, a prospectively maintained record of trainee operative experience). Assessment scores will also be explored in relation to current training year, and scores achieved in operative workplace-based assessments (SSAOP assessments). It will not be possible to relate scores achieved to subsequent patient outcomes. Correlation with tests measuring different constructs (e.g. non-technical skill assessments) will not be measured. Comparisons will be made between groups; the CST (Core Surgical Training Group) who will have undertaken simulation-based training and have gained real-world operative experience, and the MSSP (Masters in Surgical Science and Practice) cohort who have undertaken intensive simulation-based training but will not have real-world operative experience.

*Consequences:* Standard setting is conducted using a modified Angoff method; at pre-assessment rater training, consensus regarding the competence level of a 'borderline' candidate is decided upon. A global rater of 'competent', 'borderline' or 'not competent' is provided along with scores achieved in the OSATs/ checklist score. While this 'global' rater does not contribute to scores awarded, it is used to

produce a pass-fail cut-off score for each station using borderline linear regression methodology. Impact on downstream performance (e.g. downstream effects on PBA scoring, graduation rates, clinical performance or patient safety) was not measured as part of this study.

*4.3.2.3 Statistical Analysis*

Correlations between variables of interest will be explored using Spearman (continuous and ordinal/categorical variables) and Pearson (two continuous variables) coefficients. Inter-rater reliability will be assessed using intra-class coefficients. Reliability will further be examined using Generalizability theory as outlined above. Borderline regression methodology is used to generate a pass-fail cut-off score for each station by plotting total scores achieved against the global rater, and identifying the score at the line of best fit achieved by 'borderline' candidates. Statistical analysis will be performed using SPSS software (IBM, USA) and Microsoft Excel (Microsoft, USA), as outlined above under 'study outcome measures'. A p value of $<0.05$ will be considered statistically significant. A correlation coefficient of 0 to 0.19 will be regarded as very weak, 0.2 – 0.39 as weak, 0.4 – 0.59 as moderate, 0.6 – 0.79 as strong, and 0.8 – 1 as very strong[365].

**4.3.3 Results**

A total of 93 candidates completed the OSS assessment in April 2022. Of these, 76 were core surgical trainees in their first year of dedicated surgical training (CST1) while 17 were students enrolled in the Masters' of Surgical Science and Practice (MSSP) programme. Achieved scores for each procedural station and total scores by cohort are outlined in Table 4.3.1.

*Table 4.3.1: Scores for each station presented as mean ± standard deviation for each cohort (scored from 0 – 100). CST 1 indicates Core Surgical Trainees in Year 1, MSSP, Master's in Surgical Science and Practice students. Station 1 = bowel anastomosis. Station 2 = wound closure (skin) and vessel ligation. Station 3 = Abdominal wound incision and closure with drain insertion. Station 4 = Laparoscopic Basic Skills Assessment. The total score is comprised of the sum of each station score, standardised on a scale from 0-25.*

| | **Station** | | | | |
|---|---|---|---|---|---|
| **Group (N)** | **Station 1** | **Station 2** | **Station 3** | **Station 4** | **Total** |
| CST1 (76) | 82.81 ± 13.88 | 78.96 ± 12.34 | 84.87 ± 10.18 | 74.13 ± 17.04 | 80.19 ± 9.17 |
| MSSP (17) | 77.86 ± 17.68 | 68.4 ± 10.96 | 81.52 ± 11.02 | 92.14 ± 7.79 | 79.98 ± 5.80 |
| Total (93) | 81.91 ± 14.66 | 77.03 ± 12.72 | 84.26 ± 10.36 | 77.43 ± 17.21 | 80.16 ± 8.62 |

*4.3.3.1 Internal Structure Validity*

*Table 4.3.2: Internal consistency. Correlations between score domains and associated significance levels are reported across the 4 OSCE stations. OSATS indicates Objective Structured Assessment of Operative Skill score, r, Pearson correlation coefficient, rs, Spearman's rho, P, p-value. ** indicated that correlation is significant at the 0.01 level (2-tailed).*

Station 1

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Checklist | | | | |
| 2. OSATS | r .62**<br>P <.001 | | | |
| 3. Total | r .87**<br>P <.001 | r .93**<br>P <.001 | | |
| 4. Global | rs .73**<br>P <.001 | rs .72**<br>P <.001 | rs .76**<br>P <.001 | |

Station 2

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Checklist | | | | |
| 2. OSATS | r .47**<br>P <.001 | | | |
| 3. Total | r .86**<br>P <.001 | r .86**<br>P <.001 | | |
| 4. Global | rs .61**<br>P <.001 | rs .67**<br>P <.001 | rs .71**<br>P <.001 | |

Station 3

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Checklist | | | | |
| 2. OSATS | r .63**<br>P <.001 | | | |
| 3. Total | r .84**<br>P <.001 | r .95**<br>P <.001 | | |
| 4. Global | rs .56**<br>P <.001 | rs .51**<br>P <.001 | rs .57**<br>P <.001 | |

Station 4

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Checklist | | | | |
| 2. OSATS | r .84**<br>P <.001 | | | |
| 3. Total | r .96**<br>P <.001 | r .96**<br>P <.001 | | |
| 4. Global | rs .64**<br>P <.001 | rs .62<br>P <.001 | rs .65**<br>P <.001 | |

Internal consistency data is presented in Table 4.3.2. Across all skills stations, significant correlations were observed between the checklist and OSATS domains, and between these domains and the total score awarded and global competence evaluation. Correlations ranged from moderate to very strong (Table 4.3.2). Variance estimates and Generalizability Coefficients (*G*) for one individual station are outlined in Table 4.3.3.

*Table 4.3.3: Observed score variance attributable to inter-trainee variance, Var(trainee), inter-station variance, Var(Station), inter-cohort variance referring to CST vs MSSP candidates, Var(Cohort), the interaction between trainee and individual station variance, Var(trainee*station), and the residual variance attributable to error, Var(Error). G indicates generalizability coefficient scored 0-1.*

| | **Total score** | **GRS** | **Pass/ Fail** |
|---|---|---|---|
| Var(trainee) | 34.92 | 0.03 | 0.02 |
| Var(Station) | 10.64 | 0.00 | 0.00 |
| Var(Cohort) | 0.00 | 0.01 | 0.00 |
| Var(trainee*Station) | 161.04 | 0.25 | 0.11 |
| Var(Error) | 0.00 | 0.00 | 0.00 |
| G (single observation) | 0.17 | 0.16 | 0.15 |

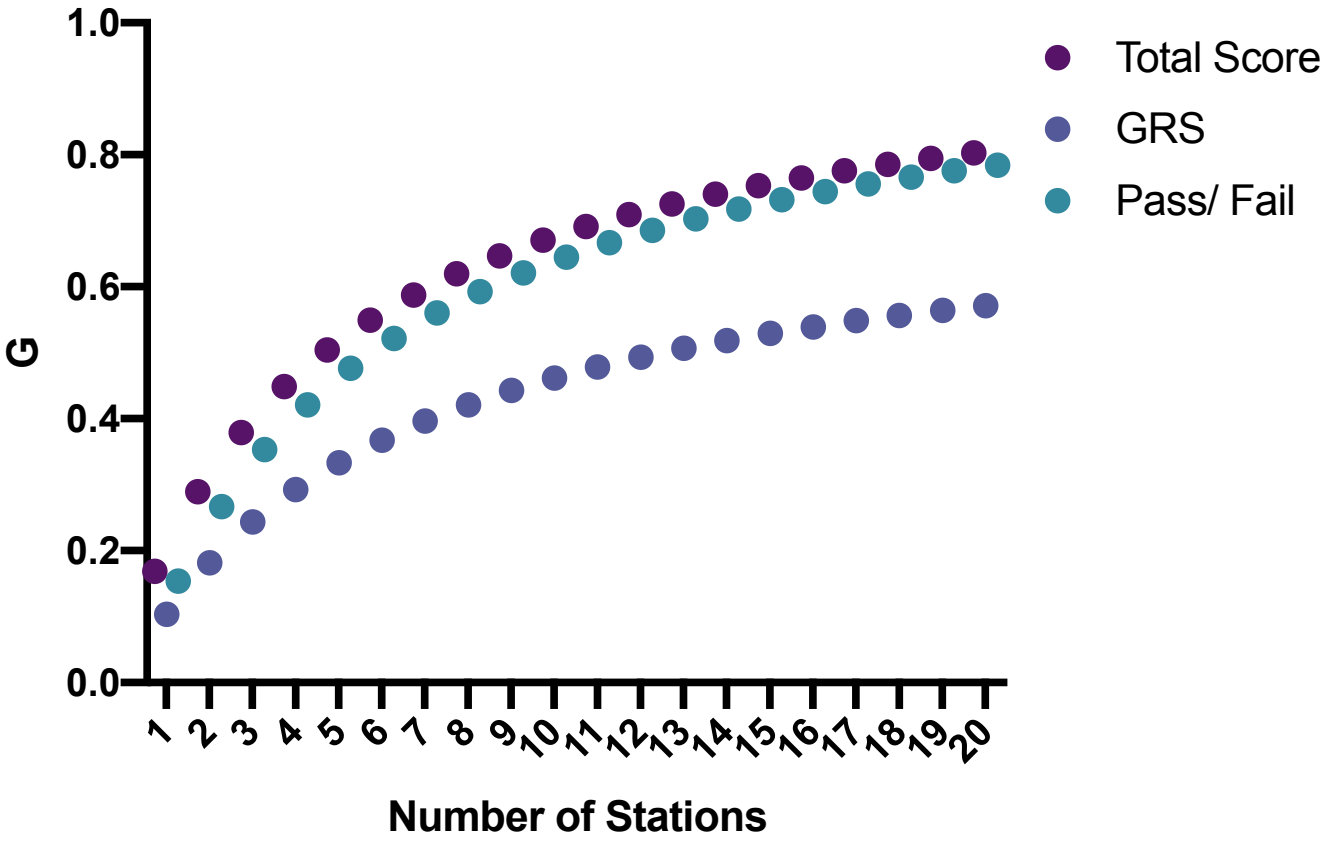



*Figure 4.3.1: Decision study. Increasing G-coefficient with increasing numbers of hypothetical stations.*

On 'Decision-study' modelling, i.e. mathematical modelling of the G-coefficient with increasing numbers of hypothetical OSCE stations, the total score (checklist and OSATS score equally weighted) was more reliable than the Global Rating Scale, and

similarly reliable when compared to the ultimate pass/ fail judgement awarded after regression analysis standard setting (Figure 4.3.1). A minimum of 12 stations/ observations scored using the total score (OSATS and Checklist) would be required to achieve a G of >0.7 (conventionally required for low-stakes assessment) and 20 observations or stations to achieve a G of >0.8 (conventionally required for high-stakes assessment). Inter-rater reliability was compared with a random sample of 25 observations (small bowel anastomosis station), scored by an independent second rater. Intra-class correlation coefficients were good for the total score (0.71) but poor for the Global Rater (0.49).

*4.3.3.2 Consequential Validity*

A pass-fail cut-off score was generated for each station using Borderline Regression Methodology (Figure 4.3.2). This resulted in a pass rate of 81.72% (76/93 candidates) for station 1, 79.57% (74/93) for station 2, 90.32% (84/93) for station 3, and 88.17% (82/93) for station 4.

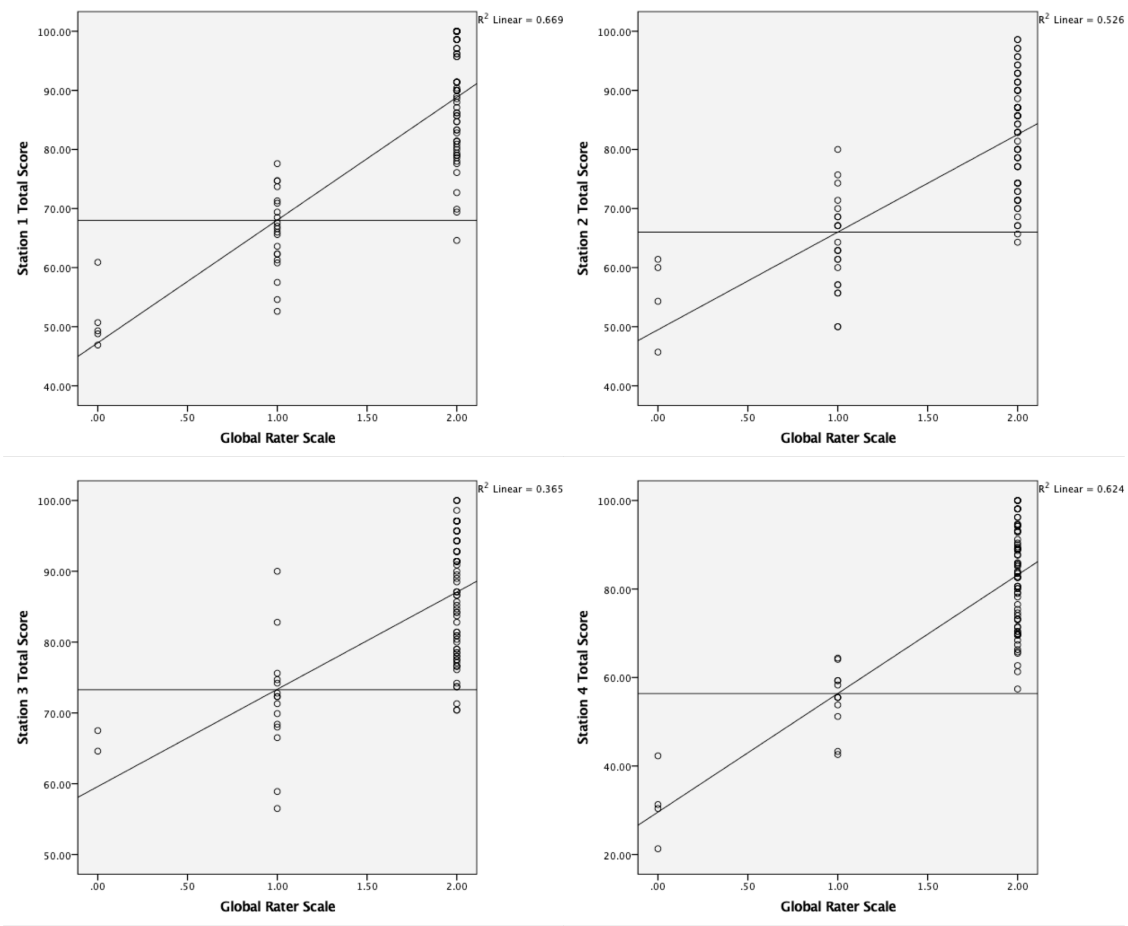



*Figure 4.3.2: Borderline Regression analyses for each station. Total scores are standardised on a scale from 0-100. On the X-axis, a GRS score of 0 indicates incompetent/ fail, 1, borderline, and 2, competent/ pass. Where the line of best fit intersects with the 'borderline' score on the X-axis, a horizontal line indicates the pass-fail cut off total score.*

### 4.3.4 Discussion

A total of 93 candidates in early surgical training (76 first year core surgical trainees and 17 Master's in Surgical Science and Practice candidates) undertook a four station operative skills OSCE. Moderate to very strong correlations were observed between score domains across stations. The inter-rater reliability was acceptable (0.71) for the total score but poor for the global competence rater (0.49). Performance across 20 stations would need to be observed for the OSS examination to be reliably used as a summative measure of operative skill for high-stakes decisions. The application of borderline-regression methodology resulted in the

production of a pass-fail benchmark of similar reliability to the total score. The station pass rate varied from 79.57% to 90.32% across stations. There was no significant difference in mean scores achieved by candidates on the MSSP programme and candidates undertaking Core Surgical Training.

The simulation-based assessments of operative performance in this study are undertaken as part of a multi-modal assessment framework. Operative skill is also assessed through workplace-based assessments of operative performance, end-of-rotation trainer evaluations and an electronic logbook which acts as a proxy measure of operative experience[401]. Further workplace-based and simulation-based assessments of non-technical skill are also undertaken during CST. The OSS assessment is undertaken at two time points; end of ST1 and end of ST2 (years 1 and 2 of CST respectively). The reliability of a 4-station OSCE in isolation is insufficient for high-stakes assessment. However, it is important to consider the assessment in the broader context of multi-modal assessment using different raters, different time points and different methods. Simulation-based assessment could compliment workplace-based assessments in surgical training. Concerns regarding the opportunistic way that workplace-based assessments are used by trainees have been raised[197 198 201 419]. Further concerns exist regarding the subjectivity of these assessments[197 198]. As they take place in the workplace, however, they are less resource intensive than simulation-based assessments and can be conducted more frequently at serial time points in order to determine trends in performance or 'learning curves'[111].

Interestingly, no difference in scores was noted between the two cohorts enrolled in this study; candidates of RCSI's MSSP (Master's in Surgical Science and Practice) programme, and the CST (Core Surgical Training) programme. MSSP and CST candidates undertake simulation-based training. MSSP candidates are full time students within the RCSI National Surgical Training programme. They complete the same surgical skills curriculum as the CST trainees, though the two year CST programme is delivered over the course of one academic year. CST trainees will undertake the simulation-based surgical skills programme over the course of two

years. As part of their clinical duties in affiliated training hospitals, they will have in-theatre operative experience. A more intensive simulation-based training curriculum delivered similar assessment performance scores to a less intensive course coupled with in-theatre training. Whether this is related to greater familiarisation with the simulation models or increased dedicated simulation teaching should be explored by future studies.

There remain significant barriers to the implementation of simulation-based assessment. Simulation is resource intensive[461]. Trainees are required to take time from their clinical duties and travel to the National Surgical Training centre in order to be assessed. There are concerns regarding the ability of current simulation technology to replicate tasks or procedures to a sufficient degree of fidelity for skills assessment[294 462]. Cost-benefit analyses of simulation-based assessment in surgery are lacking[463]. Furthermore, the impact of assessment on trainee preparation time, stress or anxiety should be further assessed. Future studies should explore this aspect of consequential validity of the OSS assessment[294].

### 4.3.5 Conclusion

This study reports validity evidence for the use of an objective simulation-based operative skills assessment in surgical training, designed to inform decisions on trainee progression. Such assessments are feasible but insufficient in determining a candidates level of operative skill to a degree of reliability facilitating high-stakes decisions. Simulation at this level should form part of a programmatic approach to trainee assessment.

## 4.4 Additional Material: Comparing the Reliability of Simulation-Based and Workplace-Based Assessments of Operative Competence in Core Surgical Training

**Abstract**

**Introduction**

Simulation-based assessments (SBAs) are more readily standardisable than workplace-based assessments (WBAs) which are prone to inter-case and inter-rater variance. This study seeks to compare the reliability of SBAs and WBAs in the assessment of operative competence of core surgical trainees in the Republic of Ireland.

**Methods**

In this prospective observational cohort study, the reliability of workplace-based Supervised Structured Assessments of Operative competence (SSAOPs) and simulation-based Objective Structured Clinical Examinations (OSCEs) were compared using generalizability theory.

**Results**

A total of 330 trainees completed 2,294 SSAOP and 995 OSCE station assessments from 2016 – 2021. Score variance attributable to individual trainee ability and procedure (or station) specific trainee ability was 12% and 57.14% for simulated OSCEs respectively, compared to 25% and 18.75% for in-theatre assessments (SSAOPS). Across repeated assessments, OSCE stations outperformed SSAOP procedural assessments in terms of reliability ($t=27.92$, $df=99$, $p<0.0001$)

**Conclusion**

Simulation is a more reliable assessment method than workplace-based assessment in evaluating the operative skill of core surgical trainees. This study supports the use of multi-modal assessment methods when measuring a trainee's operative competence.

### 4.4.1 Introduction

In Ireland, structured postgraduate surgical training begins with two years of Core Surgical Training (CST), designed to prepare trainees for competitive progression onto a higher specialist surgical training programme (ST3-ST8). Core surgical trainees generally rotate through a 6-month rotation in general surgery, a further 6-month rotation in a different surgical sub-specialty, and then a year-long rotation (ST2) in their chosen subspecialty. The operative competence of core trainees is assessed in a multi-modality fashion, using a combination of repeated workplace-based assessments, structured simulation-based OSCE assessments and trainer evaluation reports.

Workplace-based assessments of 'live' or in-theatre operative competence are performed using the Supervised Structured Assessment of Operative Performance (SSAOP) tool, which is generic (i.e. not procedure specific) and consists of a checklist score (from 0 -150) and an overall performance rater from 'unable to perform basic operative steps without direction' to 'exceptional performance, could operate independently' (Appendix 1, supplementary data). Simulation-based OSCEs (Objective Structured Clinical Examinations) are performed at the end of the first and second years of CST (ST1 and ST2). These multi-station OSCEs use a combination of synthetic and animal-tissue models to simulate procedures or part-tasks and are assessed using procedure-specific checklists developed in-house (Appendix 2, supplemental data) within the Royal College of Surgeons in Ireland[464]. A further global performance rater is scored as either 'competent', 'borderline' or 'fail'. This global rater does not affect the checklist score but is used for generating pass/fail benchmarks through regression analyses.

The reliability of a given assessment tool refers to the extent to which it reflects all possible measurements of the same construct[435]. The reliability of an assessment refers to the consistency of scores or decisions generated[465]. It is known that multiple trained assessors, on multiple occasions across multiple cases, is required to generate a reliable, objective impression of a trainee's operative ability for

summative assessment purposes[22 466]. The reliability of a single observation is not particularly useful in practice; given the inherent variability in procedures, case complexity, and assessor variability (both inter- and intra-rater) a multi-modal assessment curriculum is usually required to produce a programme of assessment robust enough upon which to base decisions on training progression[466 467]. As simulation-based assessments are more readily standardisable[468], we hypothesize that over repeated observations, OSCE stations will be more reliable than workplace-based assessments in providing an assessment of a trainee's overall operative skill.

The objective of this study is to compare the reliability of workplace-based SSAOP assessments and simulation-based OSCE assessments in the assessment of core surgical trainee operative competence using generalizability theory.

### 4.4.2 Methods

#### *4.4.2.1 Study design, setting and participants:*

Findings of this prospective observational cohort study are reported according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement 2007[349]. SSAOP assessments completed by surgical trainers for core surgical trainees from April 25th 2016 – February 26th 2021 were included. These assessments are completed through an online interface, msurgery.ie. Completed simulation-based OSCE assessments from 2016 – 2019 were also included. In-person OSCE assessments could not take place in 2020 due to the COVID-19 pandemic. Ethical approval was granted by the ethics committee of the University of Medicine and Health Sciences, Royal College of Surgeons in Ireland.

#### *4.4.2.2 Data sources/ Measurement:*

The SSAOP (Appendix 4.4.1) tool assess the operative performance of core surgical trainees in the workplace. Assessments can be initiated by trainees or

trainers, and can be undertaken for any surgical procedure which the trainer considers appropriate for the specialty and the trainee's year of training. Each SSAOP is scored by an observing surgical trainer using a 'Total Checklist' score and a global 'Overall Performance' score. The Total Checklist (TC) is scored out of a maximum of 150 marks across a number of domains: 'Case difficulty' and 'degree of prompting or direction' are scored on three-point ordinal scales, while 'Pre-operative preparation', 'Operative Procedure' and Post-Operative Care' scores are summed using a 5-point Likert scale across 6, 14 and 1 checklist item(s) respectively. The global 'Overall Performance' score is awarded on a 5-point scale from 1 - 'unable to perform basic operative steps without direction' to 5 - 'Exceptional performance, could operate independently' (Appendix 1, supplementary data). As this study is concerned with the use of the SSAOP as a measure of operative skill, only the intra-operative domain of the Total Checklist score are included in analysis.

Multi-station surgical skill OSCEs are undertaken annually during core surgical training. Each assessment consists of four stations evaluated by a trained assessor. Each station assesses a different surgical procedure or task using a high-fidelity synthetic or animal model. Stations are designed to assess core surgical skills expected of all trainees regardless of surgical subspecialty. The assessed skills are aligned with training outcomes of a two-year simulation-based training curriculum. Example stations include open varicose vein ligation, open inguinal hernia repair, vascular anastomoses, pleural drain insertion, small bowel anastomosis, and stoma formation. Assessments are scored using a procedure or task-specific checklist, with each item graded in a binary fashion as 'done' or 'not done', summed to provide a total score. Assessments are also scored using the validated OSATS scale by the same assessor[469]. The checklist and OSATs scores are equally weighted and summed to produce a total score for each station. A further 'overall competence' global rater is also provided, graded as 'Not competent/ Fail'. 'Borderline' or 'Competent/ Pass'. A modified Angoff standard setting method is used to define these benchmarks, conducted on the morning of each examination[470]. An example scoring rubric is provided in Appendix 2 (supplementary data).

*4.4.2.3 Statistical methods:*

Assessment data were anonymized prior to analysis. Generalizability theory[438] was used to analyse sources of measurement variance without the need for multiple experiments[435 439]. The reliability coefficient generated from this data can then be used to mathematically model the reliability of different test formats, for example multiple OSCE stations or multiple SSAOP assessments. Furthermore, this modelling can determine the number of assessments required to obtain a reproducible assessment of operative competence[439].

To complete the Generalizability (G) study, variance component analysis was first performed. This allows for determination of the relative contribution of individual trainees to score variance, in comparison to other sources of variance such as inter-station or inter-assessment variance and residual error. Analyses were performed for the operative domain Checklist Score of the SSAOP assessment and the procedure specific OSCE scores using the Minimum Norm Quadratic Unbiased Estimation (MINQUE) model in SPSS version 23 (IBM SPSS Statistics, Armonk, NY). A nested study design was used to determine the within-trainee effect of assessor stringency and case-to-case variation on scores; this acknowledges the fact that SSAOP assessment cases are unique to trainees[438]. The alternative - A fully crossed design - would require the same trainees to be observed performing the same cases, and is therefore not possible when real-world theatre cases are used for assessment[438].

OSCE stations and SSAOP cases were treated as random samples from an infinite universe. Variance components were used to generate a reliability coefficient (*G*) for each assessment method, where *G* = true variance/ total variance. 'True' or *intended* variance refers to the variance in scores accounted for by the individual trainee's ability. *G* is calculated on a scale from 0-1, with a value of ≥0.70 conventionally considered sufficient for low-stakes assessment (e.g. formative or non-graded assessments, or frequent assessments used to evaluate training programs), and ≥0.80 for high-stakes assessment (e.g. trainee progression or certification)[442]. Decision studies (or *D*-studies) were performed by modelling reliability analyses with

increasing numbers of OSCE stations or SSAOP procedural assessments in order to determine the minimum number of observations required to achieve a *G* of ≥0.70 or ≥0.80, conventionally deemed acceptable for low-stakes and high-stakes assessment respectively[438]. All statistical analyses were performed using SPSS software (SPSS version 23, IBM SPSS Statistics, Armonk, NY).

### 4.4.3 Results:

A total of 330 trainees submitted 2,294 SSAOP assessments, scored by 379 trainers during the study period. 180 of these trainees were also assessed across 995 simulation-based assessment OSCE stations.

*Table 4.4.1: Number of analysed assessments with mean checklist scores. SSAOP = Supervised Structured Assessment of Operative Performance (workplace-based assessment of operative performance). OSCE = Objective Structured Clinical Examination. SD = Standard Deviation.*

| **Assessment modality** | **Checklist Score (Standardised 0-1)** | |
|---|---|---|
| **SSAOP** | N | 2294 |
| | Mean | 0.842 |
| | SD | ±0.126 |
| | Range | 0.175 - 1 |
| **OSCE** | N | 995 |
| | Mean | 0.737 |
| | SD | ±0.200 |
| | Range | 0 - 1 |

#### *4.4.3.1 Relative contributions of trainee ability and procedure performed to score variance*

The relative contribution of the trainee ability (Vt), procedure being assessed (Vp) , and procedure-specific trainee ability (Vt x Vp) to observed score variance in SSAOP and OSCE assessments is outlined in Table 4.4.2.

*Table 4.4.2: Contribution of trainee ability, procedure/ station variance, and trainee procedure-specific competence to observed score variance.*

| **OSCE Assessments**<br>**Source of Score Variance** | | **Checklist score**<br>**Variance (% of total)** |
|---|---|---|
| Trainee (Vt) | Represents the contribution of individual trainee ability to observed score variance | 0.005 (12%) |
| Procedure (Vp) | Represents the contribution of individual case or station difficulty | 0.010 (24%) |
| Trainee : Procedure (Vtp) | Represents the contribution of trainee procedure-specific ability on observed sore variance | 0.024 (57.14%) |
| Error (Ve) | Residual error not accounted for by measured sources of variance | 0.003 (7.14%) |
| G (reliability coefficient) of a single SSAOP assessment | | 0.12 |
| **SSAOP Assessments**<br>**Sources of Score Variance** | | **Checklist score**<br>**Variance (% of total)** |
| Trainee (Vt) | Represents the contribution of individual trainee ability to observed score variance | 0.004 (25%) |
| Procedure (Vp) | Represents the contribution of individual case or station difficulty | 0.001 (6.25%) |
| Trainee: Procedure (Vtp) | Represents the contribution of trainee procedure- or station-specific ability on observed sore variance | 0.003 (18.75%) |
| Error (Ve) | Residual error not accounted for by measured sources of variance | 0.008 (50%) |
| G (reliability coefficient) of a single assessment | | 0.25 |

*4.4.3.2 Decision studies*

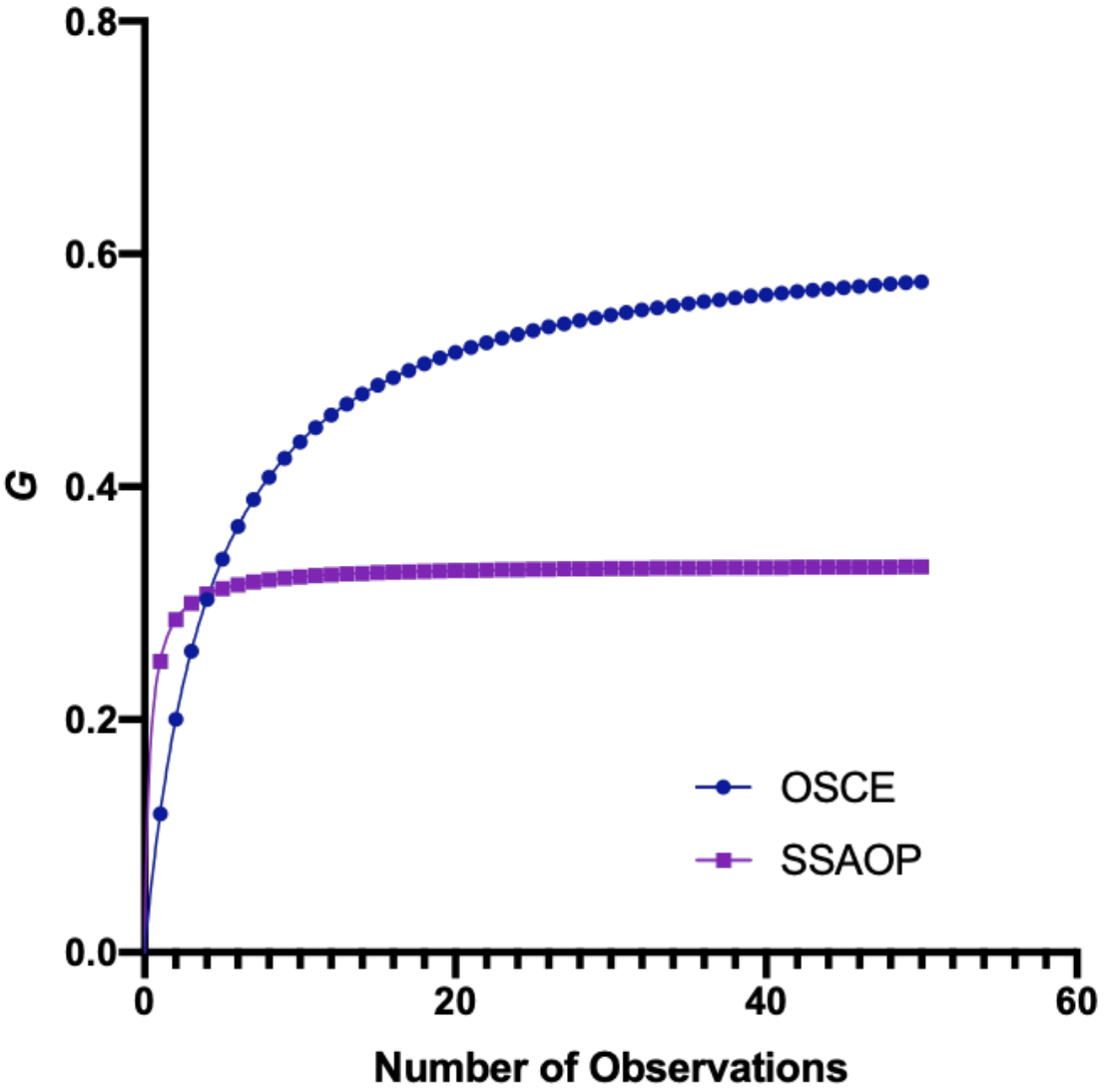



*Figure 4.4.1: Reliability co-efficient (G) of OSCEs and in-theatre (SSAOP) assessments with increasing numbers of stations (OSCE) or procedures observed (SSAOP). OSCE = Objective Structured Clinical Examination (simulation-based assessments of operative skill), while SSAOP = Supervised Structured Assessment of Operative Performance (workplace-based assessment).*

Decision studies were modelled with increasing numbers of procedural or station observations. The reliability coefficient of a single station observation (OSCE) was higher, at 0.25, than a single assessment of a given procedure in the workplace (SSAOP, G = 0.12). However, the reliability of OSCE stations were greater than SSAOP observations after ≥ 5 hypothetical procedural observations (Figure 4.4.1).

Across repeated assessments, the OSCE stations were more significantly more reliable than SSAOP procedural observations ($t=27.92$, $df=99$, $p<0.0001$).

### 4.4.4 Discussion

Over repeated observations, simulated OSCEs were a more reliable assessment of a trainee's overall operative skill than in-theatre assessments. Neither assessment modality alone is sufficiently reliable, even with increasing numbers of observations, to produce reliability coefficients suitable for making high-stake decisions such as those regarding trainee progression or certification. These findings support the use of a multi-modal assessment framework in order to assess the operative competence of surgical trainees, using multiple assessment methods scored by multiple trained assessors on repeated occasions[471]. Furthermore, these findings emphasize the importance of repeated observations, regardless of the modality, in assessing operative skill[472]. The need for multiple stations must be balanced with feasibility considerations; OSCEs are time, personnel and cost-resource intensive[473].

Simulated and in-theatre assessments can provide complimentary roles in surgical training. In-theatre assessments can provide rapid, repeated evaluations, and when used correctly are indispensable in providing targeted post-procedural feedback to trainees, and monitoring improvements in performance over the course of training[188]. Repeated observations of a given procedure can establish trainee learning curves[111], and in an ideal setting can be sufficiently reliable to evaluate a trainee's procedure-specific competence[190]. The inherent variety of in-theatre cases provides opportunities to observe a trainee's technical ability, but also non-technical skill such as judgement and decision making[474]. The SSAOP tool, for example, contains an assessment item pertaining to trainee communication, and also contains domains on pre-operative and post-operative care (though these domains were not included in the above analyses which focused on trainees' operative skill). The variety of theatre cases, however, also introduces significant variance into any workplace-based assessment, which limits the ability of these assessments to provide a 'global' evaluation of a trainee's operative skill across procedures without large numbers of

repeated observations[190 441]. With this in mind, in-theatre assessments can only be used to assess competence in procedures that are performed sufficiently frequently by trainees in practice[452].

Simulation, by contrast, is easily standardisable; trainees can be observed performing the same procedure or task on the same model, by trained assessors under the same conditions. Unlike in-theatre assessments, which must rely on the availability of given cases in order to assess a given task, simulated assessments can be tailored to assess any procedure of interest, provided that a suitable model and tool exist. For example, simulation could be used to assess performance in low-frequency, high-stakes procedures, or, as in this case, can be designed to assess performance in core skills expected of all trainees regardless of experience and intended sub-specialty. The standardisable nature of simulated-assessments has benefits in terms of reliability as outlined by this study, but also has implications for the *response-process* validity of an assessment framework; workplace-based assessments are not used in an ideal fashion, further limiting their reliability[475]. 'Gaming' of the assessment system, or the use of workplace-based assessments in an opportunistic fashion, has been described by several studies of trainees in the United Kingdom[430]. Standardised assessments with sufficient exam security measures are less prone to these practices.

Given the relative merits of each modality, a multi-faceted approach to assessment is warranted. Beyond this, given that the reliability of each modality increases with increasing numbers of observations, it is important that trainees are assessed on multiple occasions, across time points and with different trained assessors. As part of the Royal College of Surgeons in Ireland's Core Surgical Training Programme, trainees must submit a minimum of three SSAOP assessments per 6-month rotation. Simulated OSCE assessments are undertaken at the end of each training year, with four stations in each session (to a total of eight observations over two years).

*Limitations*

The SSAOP tool measures operative competence across three domains; pre-operative care, intra-operative care (procedural skill) and post-operative care. In order to produce a like-with-like comparison, only the intra-operative domain scores were included in the above analysis. The variance component analysis model used herein compares the reliability of a given station to assessments in a given procedure. There is likely significant variance in the SSAOP assessments introduced by other factors; case complexity and inter-assessor variance for example. These were not included in our Generalizability-theory (G-theory) model above, again in an attempt to compare included co-variants to those of the OSCE assessments, which are examined by a single assessor using a standardised model or 'case'. These unassessed co-variants likely account for the variance associated with 'error' outlined in the above model. It is important to note that the number of assessments required to achieve a reliable assessment does not relate to the number of procedures that a trainee needs to perform before they are competent.

### 4.4.5 Conclusion

With repeated observations, readily standardisable simulated-OSCEs are a more reliable measure of a trainee's overall operative skill. In-theatre assessments and simulated assessments have relative merits and examine different, though related, constructs. Neither modality is sufficiently reliable, even with multiple observations, to provide a summative judgement of a Core Surgical Trainee's operative competence. This study therefore supports the use of a multi-modal assessment framework, with multiple observations conducted on multiple occasions by multiple trained assessors.

## Chapter 5: Development of a novel simulation-based assessment designed to evaluate operative competence in senior general surgery trainees

This chapter outlines the process for developing a novel simulation-based assessment of operative competence for Higher Specialist Trainees in General Surgery, and subsequent pilot testing and validity evidence generation according to Messick's validity framework. Issues regarding the acceptability of a high-stakes simulation-based assessment are further explored using a series of focus groups and semi-structured interviews.

Role of the thesis author:

Mr Toale was the primary contributor to the study design and research question development for these studies, conducted the related literature searches, performed all data extraction and synthesis, drafted the initial manuscripts and produced the final versions. Furthermore, he conducted the series of semi-structured interviews and focus groups, conducted the thematic analysis of transcribed data, developed the survey items for subsequent distribution, analysed the returned data, and drafted the draft and final manuscripts.

References (published works contributing to this chapter):

Toale C, Morris M, Konge L, Nayahangan LJ, Roche A, Heskin L, Kavanagh DO. Generating a prioritised list of operative procedures for simulation-based assessment of General Surgery Trainees through consensus (Accepted for publication, Annals of Surgery)

Toale C, Morris M, Kavanagh DO. Training and assessment using the LapSim laparoscopic simulator: a scoping review of validity evidence. Surg Endosc. 2023 Mar;37(3):1658-1671. doi: 10.1007/s00464-022-09593-0. Epub 2022 Sep 19. PMID: 36123545.

Toale, C., Morris, M. & Kavanagh, D.O. Why do residents fail simulation-based assessments of operative competence? A qualitative analysis. Global Surg Educ 2, 83 (2023). https://doi.org/10.1007/s44186-023-00161-1

## 5.1 Generating a prioritised list of operative procedures for simulation-based assessment of General Surgery Trainees through consensus

Reference: Toale C, Morris M, Konge L, Nayahangan LJ, Roche A, Heskin L, Kavanagh DO. Generating a Prioritised List of Operative Procedures for Simulation-Based Assessment of General Surgery Trainees through Consensus. Ann Surg. 2023 Oct 9. doi: 10.1097/SLA.0000000000006118. Epub ahead of print. PMID: 37811854.

## 5.2 Training and Assessment using the LapSim Laparoscopic Simulator: A Scoping Review of Validity Evidence

Reference: Toale C, Morris M, Kavanagh DO. Training and assessment using the LapSim laparoscopic simulator: a scoping review of validity evidence. Surg Endosc. 2022 Sep 19. doi: 10.1007/s00464-022-09593-0. Epub ahead of print. PMID: 36123545.

## 5.3. Additional Material: Development and Validation of a Simulation-Based Assessment of Operative Competence for Higher Specialist Trainees in General Surgery

**Abstract**

**Background**

Simulation is increasingly being explored as an assessment modality. This study sought to develop and collate validity evidence for a novel simulation-based assessment of operative competence. We describe the approach to assessment design, development pilot testing and validity testing.

**Methods**

Eight procedural stations were generated using both virtual reality and bio-hybrid models. Content validity was derived from a previously conducted Delphi consensus study of trainers. Trainee performance was scored using an equally weighted Objective Structured Assessment of Technical Skills (OSATS) tool and a modified Procedure Based Assessment (PBA) tool. Validity evidence was analysed in accordance with Messick's validity framework. Both 'junior' (ST2 - ST4) and 'senior' trainees (ST 5 - ST8) were included to allow for comparative analysis.

**Results**

Thirteen trainees were assessed by ten assessors across eight stations. Inter-station reliability was high ($\alpha = 0.81$), and inter-rater reliability was acceptable (inter-class correlation coefficient 0.77). A significant difference in mean station score was observed between junior and senior trainees (44.82 vs 58.18, $p = .004$), while overall mean scores were moderately correlated with increasing training year ($rs = .74$, $p = .004$, Kendall's tau-b .57, $p = 0.009$). A pass-fail score generated using borderline regression methodology resulted in all 'senior' trainees passing and 4/6 (66.66%) of junior trainees failing the assessment.

## Conclusions

This study reports the validity, reliability and feasibility evidence for a novel simulation-based assessment, designed to assess the operative competence of higher specialist trainees in general surgery.

### 5.3.1 Introduction

The traditional Halstedian apprenticeship training model in surgery has undergone significant revolution in recent decades[1]. Evolving patient expectations regarding the role of surgical trainees in their care[2], an increased emphasis on theatre efficiency[3], and well explored concerns regarding the perceived operative competence and confidence of graduating trainees[4-6], has led to a re-evaluation of this training paradigm. 'Competency-based' approaches to outcome-driven training and assessment are well established across training jurisdictions[7 8], leading to the development of nominally time-independent postgraduate surgical programmes[9]. The surgical specialties pose a unique challenge in this regard, due to the requirement for robust and reliable methods of teaching and assessing competence in operative skill[10].

From August of 2021, surgical training in Ireland and the United Kingdom (UK) has become explicitly outcomes-based[11]. Trainees are measured against high level outcomes,(i.e. 'Capabilities in Practice'). Some trainees who demonstrate accelerated development may complete training more rapidly than the indicative time[11]. In order to transition to a truly competency-based, time-independent training paradigm, operative skill will need to be assessed using objective, standardised, and validated approaches. The primary tool for measuring operative competence remains the Procedure-Based Assessment, a workplace-based procedure-specific assessment tool[190]. Though this tool has accrued substantial validity evidence, concerns exist regarding the opportunistic way in which these assessments are undertaken by trainees, as well as their potential for rater subjectivity[559]. Operative experience targets, though a traditional proxy measure of operative competence, have been de-emphasised in the updated curriculum, and lack validity evidence in the UK and Irish context[559].

The role of simulation-based training and assessment in surgery has evolved in tandem with competency-based education[12]. Simulation is commonplace in surgical training curricula[13], and is increasingly being deployed as an assessment modality [14]

[15]. Simulation has most notably been used in high-stake assessments through the implementation of the Colorectal Objective Structured Assessment of Technical Skill (COSATS) by the American Board of Colon and Rectal Surgery[14 18]. Such simulation-based assessments have not yet been used in a high-stakes setting for higher specialist training in Ireland or the United Kingdom to date. We hypothesise that simulation could be used as a reliable, valid, objective and standardisable method of summative operative competence assessment for trainees undertaking higher specialist training in general surgery.

This study sought to develop and validate a proposed simulation-based assessment of operative competence based on the Joint Committee on Surgical Training (JCST) / Intercollegiate Surgical Curriculum Programme (ISCP) curriculum in general surgery. Herein we describe the approach to assessment design, development and pilot testing, and report on the validity of this assessment in line with Messick's validity framework[171].

### 5.3.2 Methods

#### *5.3.2.1 Initial development of the assessment framework*

The approach to assessment development and validation was derived from a primer for the validation of simulation-based educational assessment published by Cook & Hatala (2016)[173]. The primary interpretation of the intended construct was defined as: "learners have operative skill sufficient to operate safely as an independent practitioner in General Surgery". With reference to the JCST/ISCP curriculum, this construct relates to the following Capabilities in Practice (CiPs), Generic Professional Capabilities (GPCs) and syllabus requirements: 'manages an operating list' (CiP), 'professional skills' (GpC) and both 'technical skills – general' and 'technical skills – index procedures' (syllabus). The intended decision related to this examination was derived through an iterative process of literature review[500], semi-structured interviews with key stakeholder representatives[40], and a national survey of surgical trainees regarding their perceptions and experience of simulation-based assessment

(unpublished data). The scores achieved in this assessment were designed to inform, as part of a multi-faceted assessment framework, whether a given trainee has sufficient procedural competence to proceed to the next stage of training or independent practice.

This assessment was designed based on the well-established Objective Structured Assessment of Technical Skill (OSATS) examination[460] as followed by other high-stakes simulation-based assessments[14 18]. Therefore, the initial design of this assessment was that of a time-limited 8-station framework[460]. Operative procedures for use in this assessment were selected using a modified Copenhagen Academy for Medical Education and Simulation Needs Assessment Framework (CAMESNAF)[482]. This resulted in a prioritised list of procedures suitable for simulation-based assessment of general surgery trainees at the end of phase 2 (ST6), from which 8 were chosen for inclusion in the assessment model.

*5.3.2.2 Assessment Instruments*

Performance was assessed at each station using a modified Procedure-Based Assessment (PBA) tool (consisting of a task specific checklist and a global rating scale)[560] and the OSATS tool[561]. The PBA tool has accrued substantial validity evidence since its adoption by the JCST[559] for operative skill assessment. The OSATS tool is the most commonly used tool in surgical assessment research[562]. The PBA tool consists of a procedure-specific checklist containing pre-, intra- and post-operative domains, along with a global rater from 1a – 4b. Due to time and feasibility constraints, key component tasks, rather than entire procedures, were simulated for assessment in some stations. For this reason, the intra-operative domain of the PBA tool was adapted by a steering group of simulation and surgical education researchers (CT, MM, DOK). The modified PBA checklists were trialed by a consultant surgeon and surgical trainee to ensure that they could be completed on the proposed stations in the indicated time frame. Data from this trial was not included in subsequent analysis. As assessment tool modifications can have impacts

on their validity, this study also seeks to generate validity data for these modified tools.

The LapSim™ simulator was used to assess performance in three of the eight stations. This simulator reports a number of computer-measured metrics (Appendix 5.3., supplementary data). LapSim™ recorded metrics did not contribute to candidate scores, though statistical analyses are reported herein.

*5.3.2.3 Participants*

Previous literature has suggested that 8 stations provides a reliable indicator of performance in a multi-station OSCE[460]. Sample size was determined based the number of participants required to demonstrate a difference in outcomes between 'senior' and 'junior' general surgery trainees (trainees in their last or first four years of surgical training respectively). Studies reporting mean PBA scores awarded with standard deviations are lacking since the introduction of a revised global rating scale in 2017[559]. Furthermore, a novel, modified PBA tool for use in simulation-based assessment is being used in this study. For this reason, power calculations were performed using the OSATS tool. Sample size was determined based on anticipated differences in overall scores across the 8-station OSCE format, and not at the individual station level. Mean OSATS scores and standard deviations were derived from a prior study using a simulation-based assessment to assess surgeons at the end of training (mean score 0.75, SD 0.06)[563]. Assuming that a score difference of 20% or more was a relevant difference between groups, a sample size of 6 participants per assessment group was sufficient to minimize the risk of a type II error ($\alpha = 0.05$).

*5.3.2.4 Procedure*

Ethical approval was granted by the University of Medicine and Health Sciences at the Royal College of Surgeons in Ireland. The assessment took place at the National Surgical Training Centre, Royal College of Surgeons in Ireland, which contains a

purpose-built 'wet lab' for simulation-based technical skills training and assessment. Examiners were familiarized with the format, tools and models prior to the assessment, and pre-assessment standard setting was conducted using a modified Angoff method[564]. Candidates were informed of the assessment format in a 30-minute briefing session prior to commencement. This orientation session also allowed for introduction to the virtual reality models used (LapSim, Surgical Science, Sweden). Examiners were not informed of the trainee's level of training. Stations utilized a combination of bio-hybrid models (ileostomy reversal, operative management of a fistula-in-ano, pilonidal sinus excision, ventral hernia repair, emergency laparotomy and management of a blunt liver injury) and virtual reality models (laparoscopic appendicectomy, laparoscopic cholecystectomy, and right hemi-colectomy - vessel ligation). The LapSim simulator (Surgical Science, Sweden) was chosen for use in the virtual reality stations given the substantial validity evidence published with respect to its use in training and assessment[463]. Candidates rotated through the stations and were assessed by a single assessor using the modified PBA and OSATs tools. Two stations (ileostomy reversal and emergency laparotomy) were also scored by a second, independent rater. Performance in each station was video-recorded in an anonymised fashion for quality assurance purposes. Scores were entered immediately to a secure database (QPERCOM). Instructions regarding the procedure or task to be completed were given to participants before each station.

*5.3.2.5 Statistical Analysis*

Validity evidence was collated according to Messick's validity framework[171]. Internal consistency was assessed using Cronbach's alpha. Inter-rater reliability was calculated using intra-class correlation coefficients. Reliability was further explored using generalizability theory[565]; variance component analysis was used to determine the relative contribution of trainees, assessors and stations to observed score variance allowing for the estimation of the 'true' or 'intended' variance, i.e. the proportion of score variance attributable to individual trainees. For the purpose of this analysis, trainees, assessors and stations were thought of as a sample from an

infinite universe of potential trainees, assessors and stations[102]. The Consequential validity evidence was explored by generating a pass-fail cut-off score, using borderline regression methodology[566]: A linear regression model was used to plot the global competency score against both OSATS and PBA checklist scores. The global score representing the 'borderline' candidate was inserted into the linear equation, and the corresponding total OSATS and PBA checklist scores were extrapolated, representing the true pass score for each station. The pass score for the assessment as a whole was the sum of the pass scores for each station. OSATS and modified PBA scores were equally weighted. Group comparisons were performed using $\chi 2$ analysis. Relationships with other variables (training year) were performed using Pearson correlation coefficients.

### 5.3.3 Results

Thirteen candidates undertook the assessment across eight 15-minute stations. Candidates were divided into 'junior' trainees in their first 4 years of surgical training (ST1 – ST4, N = 6) and 'Senior' Trainees in their last 4 years of training (ST5 – ST8, N = 7) Mean scores awarded are outlined in Table 5.3.1.

*Table 5.3.1: Mean scores awarded across assessment stations. Station 1 – fistula in ano, 2 – ileostomy closure, 3 – right hemi-colectomy (vessel ligation), 4 – pilonidal sinus excision, 5 – ventral hernia repair, 6 – laparoscopic appendicectomy, 7 – emergency laparotomy (blunt liver trauma), 8 – laparoscopic cholecystectomy. SD indicates Standard Deviation. A competency rater score of 0 indicates a 'fail/ not competent' score, 1 indicates 'borderline' and 2 indicates a 'pass/ competent' score. PBA Global rater scores range from 0 – 8 (see supplemental data).*

| Station | Mean Standardised PBA Score (SD) | Mean OSATS (SD) | Mean Total Score (SD) | Mean PBA Global Rater (SD) | Mean competency rater (SD) |
|---|---|---|---|---|---|
| 1 | 26.65 (± 6.15) | 25.08 (± 4.39) | 51.73 (± 10.29) | 5.46 (± 0.88) | 1.62 (± 0.65) |
| 2 | 30.24 (± 6.74) | 30.76 (± 5.96) | 61.01 (± 12.41) | 6 (± 1.58) | 1.38 (± 0.65) |
| 3 | 29.72 (± 7.68) | 25.07 (± 4.65) | 54.79 (± 11.35) | 5.84 (± 1.68) | 1.77 (± 0.44) |
| 4 | 13.83 (± 11.59) | 23.54 (± 7.86) | 37.37 (14.64) | 5.69 (± 1.75) | 1.00 (± 1.00) |
| 5 | 26.56 (± 7.33) | 26.17 (± 3.07) | 52.73 (± 10.25) | 5.42 (± 1.24) | 1.42 (± 0.90) |
| 6 | 27.12 (± 12.85) | 30.69 (± 4.21) | 57.82 (± 15.21) | 5.62 (± 1.66) | 1.69 (± 0.63) |
| 7 | 17.5 (± 8.36) | 21.31 (± 3.95) | 38.81 (± 12.21) | 3.54 (± 2.29) | 0.46 (± 0.78) |
| 8 | 31.41 (± 5.52) | 32 (± 4.51) | 63.41 (± 9.02) | 6.83 (± 1.80_ | 1.92 (± 0.29) |

*5.3.3.1 Content Validity*

Processes for the development of stations, examination content, score checklists and models are outlined above under 'Methods'.

*5.3.3.2 Internal structure validity*

Inter-station reliability for the total score awarded was assessed using Cronbach's alpha ($\alpha$), calculated as 0.81 for the assessment overall. Removing station 1(management of a fistula-in-ano) would result in a higher $\alpha$ (Table 5.3.2).

*Table 5.3.2: Cronbach's alpha. PBA indicates Procedure-Based Assessment tool, OSATS, Objective Structured Assessment of Technical Skill.*

| **Score** | $\alpha$ |
|---|---|
| PBA | 0.68 |
| OSATS | 0.85 |
| Total Score | 0.81 |
| PBA global rater | 0.81 |
| Competency rater | 0.71 |

Internal consistency of individual assessment tools is outlined in Appendix 5.3.2, supplementary data. Correlations between individual station scores and the total score, and between scores awarded at each station, are reported in Appendix 5.3.3 and Appendix 5.3.4 (supplementary data).

Inter-rater reliability was assessed for two stations: emergency laparotomy (Station 7) and ileostomy reversal (Station 2). On intra-class correlation coefficient analysis, inter-rater reliability was 0.77. Individual station scores variably correlated with the total score awarded (Table 5.3.3). Using Generalizability Theory, a reliability coefficient (G) of 0.7 was calculated for the overall 8-station assessment. Fourteen hypothetical stations would be required to increase the reliability coefficient to >0.8. Further data relating to the model construction and variance component analysis is outlined in Table 5.3.3.

*Table 5.3.3: Correlation (Pearson) between individual station scores and total scores awarded. r indicates Pearson correlation coefficient, p, p-value, *, p<0.05, **, p<0.005, ***, p<0.0005*

| **Station** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
|---|---|---|---|---|---|---|---|---|
| **r** | .48 | .66 | .63 | .69 | .54 | .89 | .84 | .57 |
| **p** | .10 | .02 * | .02 * | .01 * | .07 * | <.001 *** | <.001 *** | .06 |

*5.3.3.3 Relationships with other variables.*

A significant difference in mean station score was observed between 'Junior' and 'Senior' trainees (44.82 vs 58.18, p = .004) (Figure 5.3.1). Mean scores were moderately correlated with increasing training year (rs = .74, p = .004, Kendall's tau-b .57, p = 0.009) (Figure 5.3.2).

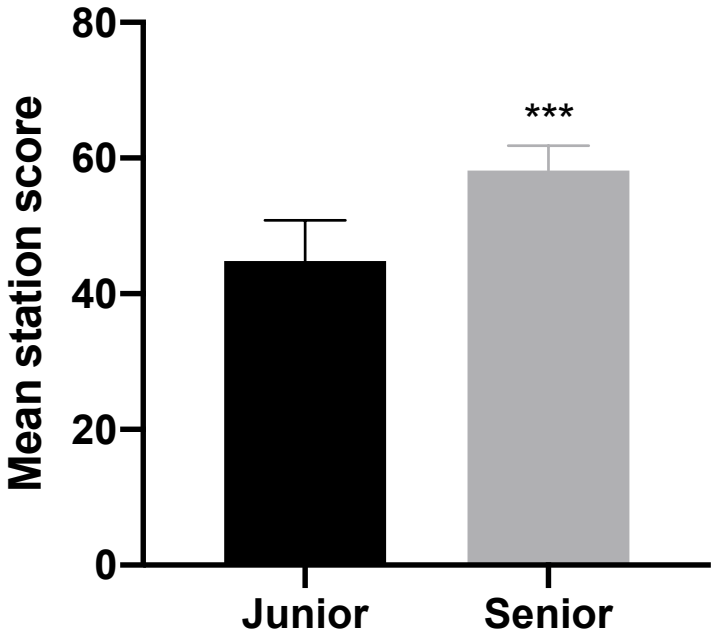



*Figure 5.3.1: Difference in mean scores achieved by 'Junior' (ST2 – ST4) trainees, and 'Senior' (ST5 – ST8) trainees.*
**** indicates p<0.005*

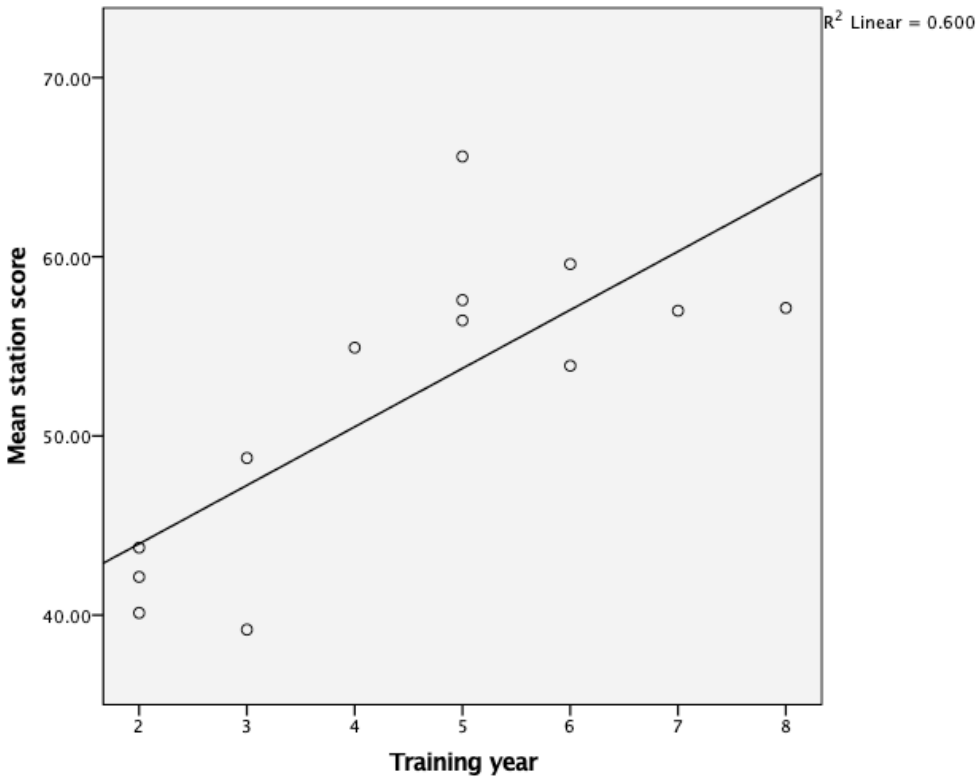



*Figure 5.3.2: Relationship between training year and mean score obtained*

At an individual station level, five of eight stations could differentiate between 'Junior' and 'Senior' trainee candidates (Table 5.3.4).

*Table 5.3.4: Mean score differences for 'Junior' vs. 'Senior' trainees across individual stations.*

| Station | Group | Mean | Std. Deviation | Sig. (2-tailed) |
|---|---|---|---|---|
| 1. Fistula-in-ano | Junior | 45.4722 | 9.12105 | **.035** |
| | Senior | 57.0952 | 8.35378 | |
| 2. Ileostomy closure | Junior | 52.6944 | 14.47427 | **.017** |
| | Senior | 68.1429 | 2.24522 | |
| 3. Right hemi-colectomy | Junior | 47.4038 | 12.21871 | **.022** |
| | Senior | 61.1319 | 5.63641 | |
| 4. Pilonidal sinus excision | Junior | 32.0530 | 10.95141 | .241 |
| | Senior | 41.9221 | 16.62621 | |
| 5. Ventral hernia repair | Junior | 49.9167 | 10.40032 | .367 |
| | Senior | 55.5417 | 10.19978 | |
| 6. Laparoscopic appendicectomy | Junior | 46.4551 | 15.99882 | **.006** |
| | Senior | 67.5659 | 3.11505 | |
| 7. Emergency laparotomy (blunt liver trauma) | Junior | 28.2083 | 4.13017 | **.000** |
| | Senior | 47.8929 | 8.70474 | |
| 8. Laparoscopic cholecystectomy | Junior | 59.7538 | 10.96857 | .254 |
| | Senior | 66.0220 | 7.06335 | |

A number of simulator-measured metrics were further recorded for the three laparoscopic skills (LapSim™) stations ( Appendix 5.3.1, supplementary data). Correlations between simulator- and assessor- recorded metrics are outlined in Appendices 5.3.5, 5.3.6 and 5.3.7 (supplementary data). Performance metrics as measured by the LapSim were not consistently associated with performance as scored by trained assessors. Similarly, LapSim metrics did not consistently discriminate between junior and senior trainees.

*5.3.3.4 Response process validity*

Video capture was used for quality control purposes. Rater training was conducted before the assessment. All assessors were registered consultant surgeon trainers with the Royal College of Surgeons in Ireland.

*5.3.3.5 Consequences*

A pass-fail standard for each station was generated using borderline regression methodology (Figure 5.3.3).

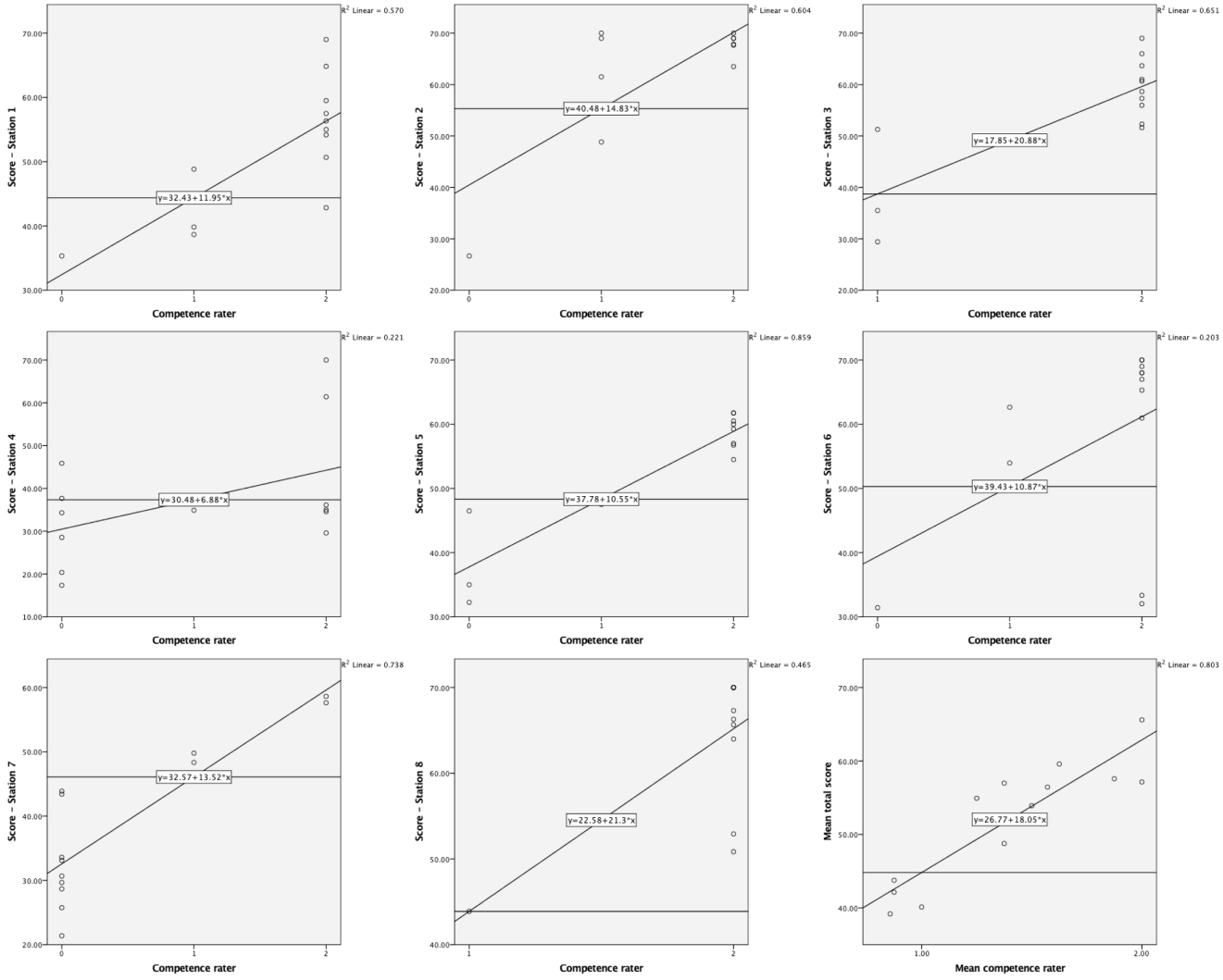



*Figure 5.3.3: Generation of a pass-fail benchmark*

Passing scores and passing rates for individual stations are outlined in Table 5.3.5.

*Table 5.3.5: Borderline regression passing scores and passing rates. SD indicates Standard Deviation*

| Station | Mean score (SD) | Passing score | Passing rate (%) |
|---|---|---|---|
| 1 – Fistula-in-ano | 51.73 (± 10.29) | 44.38 | 69.23 |
| 2 – Ileostomy closure | 61.01 (± 12.41) | 55.31 | 76.92 |
| 3 – Right hemi-colectomy | 54.79 (± 11.35) | 38.73 | 84.62 |
| 4 – Pilonidal sinus excision | 37.37 (± 14.64) | 37.36 | 30.77 |
| 5 – Ventral hernia repair | 52.73 (± 10.25) | 48.33 | 66.67 |
| 6 – Laparoscopic appendicectomy | 57.82 (± 15.21) | 50.30 | 76.92 |
| 7 – Emergency laparotomy | 38.81 (± 12.21) | 46.09 | 30.77 |
| 8 – Laparoscopic cholecystectomy | 63.41 (± 9.02) | 43.88 | 100 |
| Total Score | 52.01 (± 8.36) | 44.82 | 69.23 |

There was no difference in the pass rate (9/13, 69.23%) when using a compensatory model (where the candidate can fail an unlimited number of stations once the overall pass mark is achieved) or a conjunctive model (where the candidate is required to achieve the overall pass mark and pass at least 50% of stations). Either method resulted in all 'senior' trainees passing, and 2/6 (33.33%) of 'junior' trainees passing. Only one trainee passed all eight stations.

*5.3.3.6 Feasibility:*

Two rounds of assessment were conducted. Each session was conducted over 2 hours. Models and required equipment for open stations cost an estimated €4,680, though much of the reusable equipment was already available within the simulation laboratory at RCSI. Rental and support costs for the LapSim laparoscopic simulators were €4,305. Staffing costs were estimated at an additional €4,000, bringing the estimated cost of the assessment to €12,985.

### 5.3.4 Discussion

This study outlines the feasibility, reliability and validity of a novel simulation-based assessment of operative competence, designed for assessment of higher specialist trainees in General surgery and centered around operative competency expectations derived from the intercollegiate surgical curriculum programme of Ireland and the United Kingdom. High internal consistency metrics were generated,

with acceptable inter-rater reliability. While a sufficient number of observations were observed to reliably rank candidates (G >0.8 with 6 stations), 14 stations would be required in order to reliably determine whether a candidate met the desired level of operative competence across procedures, due to the observed absolute error variance[102]. The assessment can differentiate between trainees according to level of training, with more senior trainees outperforming their junior trainee counterparts across management of a fistula in-ano, ileostomy closure, right hemi-colectomy, laparoscopic appendicectomy and emergency laparotomy stations. Mean station scores were moderately and significantly correlated with increasing training year ($R^2$ = 0.6). Using borderline regression methodology, a pass-fail mark was generated which resulted in all trainees in their last 4 years of training passing the assessment, and 4/6 more junior trainees failing. The estimated cost of delivering this pilot assessment was €12,985.

The implementation of competency-based education principles across surgical training programmes requires robust methods of trainee assessments. The newly updated general surgery curriculum of the United Kingdom and Ireland emphasizes two critical stages during higher specialist training with corresponding operative competency expectations; phase 2 outlining expectation in core elective and emergency procedures, and phase 3 outlining further competency expectations in a given trainee's sub-specialty of interest. Assessment across these procedures is primarily conducted in the workplace. In other jurisdictions, high-stakes simulation-based assessments have been designed to inform end-of-training certification decisions[14 15 18 567]. This study suggests that a similar assessment modality has sufficient validity evidence to support its use in the assessment of trainees in Ireland. Its role, if any, in informing high-stakes training decisions such as progression or end-of-training certification, however, remains to be fully elucidated, particularly in the context of an existing longitudinal multi-faceted assessment programme.

In a 2012 Delphi study of members of the Association for Surgical Education , clarifying the role of simulation for the certification of residents and practicing surgeons was identified as a high-priority for surgical education research[91].

Simulation-based assessments (SBA) currently plays varying, though often significant, roles in the training and certification process across jurisdictions. For example, passing the fundamentals of laparoscopic surgery assessment is a certification requirement in certain specialties in the United States[72-76]. At end-of training certification level, studies to date suggest that SBA assesses a different construct to knowledge-based assessments, and may therefore add validity to the certification process[18 83]; a low positive correlation only (r = 0.25) was observed between scores awarded to trainees in a simulation-based certification exam and oral examination scores by de Montbrun et al., 2016[18]. Despite published validity evidence for such assessments according to modern concepts of construct validity, the acceptability of such assessments remains a barrier to their widespread implementation. Stakeholders report lack of evidence and financial considerations as factors contributing to lack of SBA implementation[37].

These concerns, particularly with respect to cost, are not unfounded. The estimated costs of the reported assessment approached €13,000. This assessment took place in a well-equipped national training center with purpose-designed facilities and trained personnel. The implementation of such assessments across programmes and local contexts will therefore likely vary. Furthermore, studies have yet to determine the appropriate role and weighting of simulation-based assessments in the context of programmatic assessment. Potential drawbacks of SBA when compared to workplace-based assessment include insufficient model fidelity, difficulty in simulating case complexity or variation, and further challenges associated with conducting longitudinal, repeated assessments. Finally, the acceptability of such assessments to key stakeholders in surgical education should be further explored by future studies prior to the implementation of any high-stakes assessment.

*Limitations:*

This study is limited by its small sample size. Derived procedures, assessment tools and competency expectations may differ across jurisdictions, training stages and local contexts. The somewhat unique highly centralised, nationally delivered and relatively well-resourced simulation-based training and assessment curriculum of Ireland may mean that further implementation challenges would be encountered by other jurisdictions.

### 5.3.5 Conclusions:

This study reports the validity, reliability and feasibility evidence of a novel simulation-based assessment of operative competence, designed for assessment of Irish higher specialist trainees in General surgery. Further larger scale validation studies, along with studies exploring the role of such assessments in a longitudinal training and

## 5.4. Additional Material Using simulation to assess operative competence: a qualitative case study of a novel simulation-based assessment

**Abstract:**

**Introduction**

Simulation is being increasingly explored as a modality for high-stakes assessment in medical education and training. The extent to which simulation can be used to assess real-world operative performance is debated, with trainers and trainees remaining skeptical regarding its use in summative settings. This study used a novel simulation-based assessment (SBA) of operative competence to explore the extent to which simulation can and should be used to assess operative competence in surgical training.

**Methods**

This qualitative case study was informed by a post-positive, social constructionist paradigm. The available relevant surgical literature, prior conducted qualitative research and participant observation was used to inform the research focus, study design and interview tool development. Group and individual semi-structured interviews were conducted with a purposive sample of 10 trainees and 10 surgical trainers who took part in a novel simulation-based assessment of operative competence. Reflexive thematic analysis was conducted using an inductive and constructivist framework. Initial open coding was followed by secondary deductive coding informed by Ten Cate's conceptualization of medical competence as a multi-layered construct.

**Results**

Trainers and trainees regard simulation-based assessment as having a primary role in determining canonical knowledge and skill, best employed prior to real-world operating as part of a train-to-proficiency program. Though this is more suited to early surgical training, SBA may be used in later training to ensure competence prior to exposure to new procedures, or as part of a programmatic certification process,

assessing procedures too infrequently encountered in the practice to rely on workplace-based assessment alone. The assessment of context-dependent knowledge is more acceptably performed in the workplace, though simulation of high physical and psychological fidelity may bridge this gap. ‘Personalized competence’ or ‘the art of medicine’ cannot be meaningfully assessed using simulation.

## Conclusion

Simulation-based assessment can be most acceptably used to ensure a minimum standard of canonical knowledge and skill prior to undertaking real-world operating in a procedure-specific fashion, with a view to ensuring patient safety, improving the efficiency of surgical training and improving trainee confidence. Single time-point, high stakes assessments of operative competence in commonly performed procedures are unlikely to be able to assess context-dependent knowledge required for certification purposes.

### 5.4.1. Introduction

Competency-based education principles have been embraced by surgical training programmes across jurisdictions[568-571]. In Ireland and the United Kingdom, 2021 saw the introduction of a competency-based curriculum, driven by the assessment of trainee outcome abilities. As Ten Cate et al. (2023) noted however, attempts at reaching a definition of 'competence' in medical education are often perceived as reductionist, failing to capture the broad and holistic construct of competent performance[572]. As it related to the performance of operative procedures, competence may encompass procedural and anatomical knowledge, technical skill and dexterity, clinical reasoning, communication, inter-disciplinary teamwork and leadership, among other technical and non-technical skills. Reliable and valid assessment of such a multi-faceted construct likely requires multiple assessment modalities, as part of a longitudinal, programmatic approach.

Simulation has been explored as an operative skill assessment modality, most notably in a number of high-stakes contexts such as end-of-training certification[14 15 567] and progression through training[44]. The extent to which simulation-based assessment can measure operative competence is essential in determining its future role in the assessment of surgical trainees. Research to-date suggests that stakeholders remain skeptical regarding its use in summative contexts[294].

This qualitative study sought to explore the extent to which simulation-based assessment can acceptably be used to assess operative competence, with a view to determining its role in competency-based surgical training curricula. A case study of a novel simulation-based assessment of operative competence, designed for use in the assessment of Irish Higher Specialist Trainees in General Surgery, was conducted in order to explore the experiences of surgical trainers and trainees with regard to simulation-based assessment.

### 5.4.2 Methods:

This study is reported according to the COnsolidated criteria for REporting Qualitative studies (COREQ) guidelines 2007[573].

#### *5.4.2.1 Context*

Surgical training in Ireland is guided by the Intercollegiate Surgical Curriculum of the UK and Ireland; trainees undertake two years in Core Surgical Training (CST), followed by six years of Higher Specialist Training (HST) in the trainee's sub-specialty of interest. Assessment of operative competence is undertaken in a multi-modal, programmatic fashion using primarily workplace-based assessments (WBAs). Simulation is not formally utilized as an assessment modality at HST level, though simulated procedural task assessments do at present inform decisions on competitive progression from CST to HST in Ireland. This case study forms part of a larger mixed-methods study exploring the utility of simulation as a high-stakes assessment modality in surgical training. A novel simulation-based assessment was designed to assess trainees at the end of 'phase 2' of surgical training; at this juncture trainees are expected to be competent in performing a series of core elective and emergency general surgery procedures, to the level expected of a day-1 consultant. An eight-station Objective Structured Clinical Examination was developed, assessing performance across eight procedural stations using a combination of bio-hybrid and virtual reality models. The novel assessment took place in the Royal College of Surgeons in Ireland's National Surgical Training Centre, a well-resourced and purpose-built simulation facility with a faculty of trained simulation technicians and administrative personnel.

#### *5.4.2.2 Study Design*

This case study was informed by a post-positivist, social constructionist paradigm. A post-positivist approach recognizes an objective truth but values subjectivity and multiple stakeholder perspectives, further acknowledging the motivations and

existing knowledge of the researcher as informing the research process and outcome[574]. Social constructionism posits that how we understand and perceive reality is context dependent, relying on the culture and times we live in[575]. Participants were recruited through purposive sampling; surgical trainers and trainees undertaking a novel simulation-based assessment (SBA) of operative competence. Participants were approached face-to-face and via e-mail. Ten surgical trainees and ten surgical trainers took part in this study, across group and individual semi-structured interviews. Four surgical trainees who took part in the SBA did not participate in further group or individual interviews. Group interviews took place within the Royal College of Surgeons in Ireland's National Surgical Training center, while individual semi-structured interviews were conducted using online video-conference software. Only researcher and interviewees were present. Recruits were consultant trainers in general surgery and surgical trainees, ranging in experience from ST2 (in their second year of dedicated surgical training) to ST8 (trainees in their last year of training). Interviews were recorded using LearningSpace™ (CAE Healthcare, International) software (for in-person interviews) and online video-conferencing software (Microsoft Teams, Microsoft USA) for online individual semi-structured interviews. Audio files were transcribed using automated transcription software (NVIVO 12) and verified by the first author (CT). Data was supplemented by field notes taken during the time of interview, and during observation of trainers and trainees participating in the SBA. Interviews were conducted from August 29th 2022 – August 15th 2023, with group interviews taking place on the day of the assessment and further individual interviews taking place in a sequential fashion. Group interviews occurred simultaneously; MM interviewed consultant trainers while CT interviewed participating trainees. The same interview tools were used by both researchers, and field notes compared. Interviews ranged in length from 27 minutes and 12 seconds to 45 minutes and 51 seconds. Interviews were analysed sequentially until data sufficiency was reached. An interview guide was developed through an iterative process of literature review, piloting, and prior qualitative research[294 329]. Transcripts and findings were returned to participants for comments and correction.

*5.4.2.3 Reflexivity*

CT is a surgical trainee with a research interest in surgical education, assessment and simulation. CT has experience of simulation as both an assessment and training methodology, having completed RCSI's Core Surgical Training program which uses SBA to inform progression through training. As such, he had a pre-existing working relationship with some of the participants of this study, and his own biases, assumptions and experiences of simulation likely inform the analysis herein. MM is a senior lecturer with research expertise in surgical education and assessment, while DOK is a consultant surgeon, trainee supervisor and surgical education researcher. We have therefore considered our differing experiences and perspectives on simulation as an assessment modality and have sought to proffer counterpoints and challenge assumptions during data collection and analysis.

*5.4.2.4 Analysis and findings*

Reflexive thematic analysis was used to analyse the data using a constructivist framework. Data was initially coded by a single researcher (CT). Definitive codes, themes and sub-themes were discussed and agreed by the research team (CT, MM, DOK). A description of the coding tree is provided within the results section. Initial coding was open and inductive. Further or secondary coding was informed by Ten Cate et al.'s conceptualization of medical competence as a multi-layered construct[572]. This describes competence according to three layers; canonical knowledge and skills (layer 1), context-dependent medical competence (layer 2) and 'personalized' competence (thought of as 'the art of medicine', (layer 3)[572]. NVIVO software (NVIVO 12, QSR International) was used to manage the data.

### 5.4.3 Results

Ten surgical trainers and eleven surgical trainees participated across group and individual interviews. Two of the ten surgical trainers took part in subsequent individual semi-structured interviews. Two surgical trainees did not take part in the initial group interviews but subsequently participated in individual semi-structured interviews. Participants included a range of trainees of different levels of seniority, from ST2 to ST8. All transcripts were used in the resulting analysis. The generated themes, informed by Ten Cate et al.'s conceptualization of medical competence as a multi-layered construct, are outlined below (Figure 5.4.1).

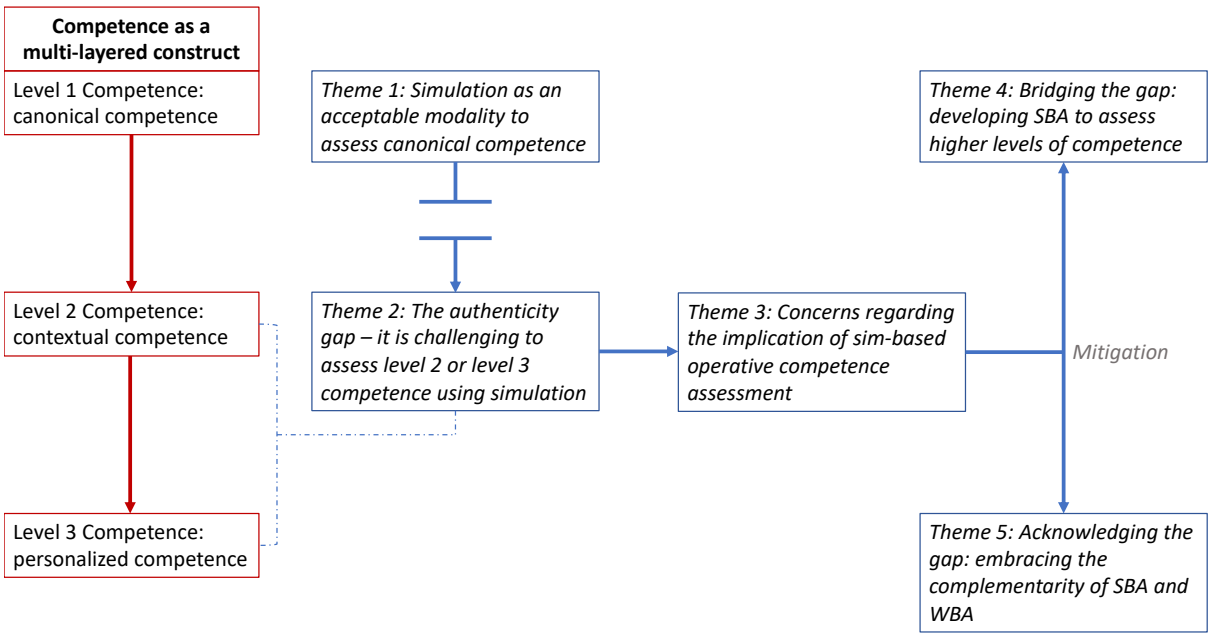



*Figure 5.4.1: Generated themes*

*5.4.3.1 Simulation as an acceptable modality to assess canonical competence*

Canonical skills and knowledge – described by trainers and trainees as anatomical and procedural knowledge, knowing 'the steps' of an operation, and basic surgical skills, are readily accepted by trainers and trainees alike to be the foundational levels of competence, essential to be obtained prior to developing context-dependent competencies and skills: *"I think at the most basic level you should know the anatomy. Then after that the next level up is knowing the steps of the operation. Those two kind of go hand in hand and those are things you can learn sitting at home reading. Then it goes into performing the operation and the actual finer technical skill would probably be the next level up from that" – Trainee*. *"The steps of the operation are the easy part, it's the stuff around the operation that's the tricky part" – Trainee*

Experienced simulation-based assessments are consistently reported by trainers and trainees as having sufficient professional authenticity to asses these basic, canonical competencies; this relates to the perceived value of SBA at more junior levels of surgical training: *"I think it is good for that exact purpose, like getting through the steps of a procedure, you know, from start to finish [...] So maybe for earlier years, it's like to be able to get through your full procedure from start to finish with all of the sequential steps? It's good for that". – Trainee*

Participants largely support the idea of simulation being used to assess to a 'minimum standard' of canonical competence in core procedures, prior to undertaking in-theatre operating as the primary, supervised, surgeon. This has perceived benefits for patient safety and training efficiency: *"It's a bridge to doing the operation in real life. If you don't cross that bridge, I think, doing it […] in real life will become, you know, exceedingly more difficult". – Trainer*

Trainees relate this concept to existing robotic training courses which mandate reaching pre-defined competency metrics prior to undertaking in-theatre supervised operating: *"I think like the robotic platform is a really nice example of that, like that's*

*how everyone learns - they're on the simulator and they do their 50 hours or whatever their, you know, country recommends they do before being allowed to start up their robotic practice" – Trainee*

*5.4.3.2 The authenticity gap – higher levels of operative competence are better assessed in the workplace*

Participants perceive operative competence to be a broad construct incorporating a wide array of technical and non-technical skills, and report context-dependent and personalised skills to be greater differentiators of competence than procedural or anatomical knowledge at more senior levels of training: *"I'm not really sure [SBA] is a reliable way to, you know, say that that person is ready to, you know, begin consulting practise […] There's so many other elements to it that it's not assessing like communication and higher level thinking and troubleshooting - "And if this doesn't work, what are you going to do next?" - and those aspects which are really what distinguishes […] someone who's just able to get through an operation for someone who's ready to take that next step in their career." – Trainee.* In particular, the ability to apply one's skills and knowledge to unexpected events or high-pressure scenarios was reported to be a marker of higher competence: *"I think ultimately […] somebody at the end of their training has to have been exposed to, you know, all of these nuances […]They need to be deemed competent, because when they do become a consultant it will be them, and They have to then be able to show maturity and leadership, and be able to handle such, you know, unpredictable situations that arise sometimes in surgery." - Trainer*

It is felt to be more difficult to assess senior trainees as SBA is perceived as not capturing these competencies, particularly as SBA cannot perfectly replicate real world case complexity or the operating environment: *"I guess, kind of, the 'simulation artefact', or whatever the technical term is when you're used to doing cases in real life… I think the gap is just too… too wide for it to be believable, I think, at a higher level, and the kind of… the nuances of what's actually difficult in the case […] it just felt so far removed from what it's like to actually manage tissue."*

More experienced trainees may be more aware of the differences between simulated and patient-operating. Comparisons to workplace-based assessment were common, and emphasised the authenticity of these assessments in determining competence in context-dependent skills, and beyond: *"you know, communication and leadership […] and management and self-awareness and all this sort of stuff… I mean, that… that's better served in a real life scenario than in a simulated exercise, you know?" – Trainer*. Longitudinal assessment *'by more than one trainer, across a period of time'* is desirable [Trainee], with the value of SBA beyond expert judgement in the workplace questioned: *"I don't think you could say [a trainee has], you know, done a hundred lap choles, his assessors have signed him off as being, you know, appropriate for stage. And now we brought him in, and we looked at him doing it on a pig model, and we think he's not safe?" – Trainee."* A one-off SBA was reported to be unlikely to capture real-world case variety or complexity.

Many of these listed competencies – communication, leadership and stress management – require an operative environment with which to interact; simulation may struggle to replicate this complex system sufficiently to assess competence in these domains: "I think it's something that's very difficult to replicate exactly - the stress surrounding it, the noise, you know? Like you having to focus on your bit of the operation and let everybody else do their other bit. But at the same time, you kind of need to know and be aware of everything that's going on in the room". Furthermore, due to differences between sim-models and in-theatre operating, there is felt to be a learning curve associated with the model itself, contributing to a perception that SBA tests model-related competence only – *"have they gotten better at doing the model or […] have they gotten better at, you know, doing the operation?" - Trainee*

Participants report that this authenticity gap leads to in-authentic trainee performance, with trainees adapting their workplace behaviour to suit the limitations of the assessment. This was reported primarily by more senior trainees: *"I wasn't doing what I normally do, I was doing what I thought I needed to do to achieve the task."- Trainee.* In fact, the very nature of the assessment environment means that

trainees adapt their behaviour to conform to what they perceive as being likely to score highly in assessment: *"It's very easy to just slip in to 'Oh yeah I'm in RCSI […] and I would do that and I would do this' and it's like, a lot of times it's not what you actually do on the day, you're saying 'oh I'll probably use that prolene because I saw someone do that before me', but I would never do that if I was doing it on a real patient."* Naturalistic performance, as it occurs in the workplace, may therefore not be assessed in the simulated environment, *"Particularly whenever your model isn't bleeding catastrophically in front of your eyes." – Trainee.* While some of this authenticity gap is introduced by imperfect authenticity of the models and assessment environment, some is due to inherent limitations of SBA, such as time constraints: *"I was like - I would obviously pack all four quadrants and do a 10 minute [pause]... but I'm not going to do that here."- Trainee.*

*5.4.3.3 Concerns regarding the implications of sim-based operative competence assessment*

The perceived inability of simulation to capture the construct of competence lead trainees and trainers to report concerns regarding the implications of SBA should it be introduced to assess competence. These concerns related primarily to the idea that SBA could be used to supersede trainer judgement of workplace performance, and could unfairly hamper the progress of trainees through training: *"Even if you have a perfect score in the college but they think you're unsafe, you know - their decision overrides*" – Trainee. Further concerns related to the concept that inauthentic assessment could lead to negative skill transfer to the workplace: *"There were certain things that I had to do to complete [the simulated procedure] that I would consider unsafe" – Trainee.*

*5.4.3.4 Bridging the gap: developing simulation-based assessment to assess operative competence*

Recognizing a difference in authenticity between assessment in the workplace and in the simulation suite, and relating this to an inability of simulation to capture higher levels of competence, respondents suggest that SBA should be developed and improved in order to approach near-real-world levels of professional authenticity. Participants were optimistic regarding the ability of future iterations of SBA to better replicate real-world operating, and report examples of sim-based training courses that simulate higher levels of psychological and physical fidelity: "It exists right? Cause like we've all done those MASCOT days where we're there with anaesthesia […] they'll just hit a button [on the simulator] and they'll make the airway extraordinarily difficult for them." Within the context of the novel simulation-based assessment experienced by respondents, some stations were felt to be better discriminators of higher-level competence than others, and approached in-theatre levels of authenticity: *"I felt that at the end of the day I had done a full operating list" – Trainee.*

Beyond improving assessment in line with best practice, it was reported that scenarios could be designed to assess context-dependent competence by probing skills such as operative planning and responding to case-to-case variation. Increasing physical and psychological fidelity of the assessment was reported to be more likely to capture higher-level competencies, and models should therefore attempt to replicate the real-world operating environment to a high degree: *"If you put it in a situation where, I don't know, there were monitors beeping in the background so you added an audio element to it… you could have introduced an anesthetist as well, a scrub nurse to also, kind of, simulate the steps of the actual procedure […] I mean, you're making a bit more high fidelity and in situation and possibly you could say, "OK, I mean maybe they can overcome some high pressure situations or maybe they can't" – Trainee.* The closer the simulator or simulation approached real life, the lesser the perceived model-related learning curve – *"the better the fidelity, the less of a, kind of, a difference in the learning curve"- Trainee.*

5.4.3.5 Acknowledging the gap: embracing the complementarity of workplace-based assessment and simulation-based assessment

Workplace-based assessment has professional authenticity perceived to be difficult to reach using simulation. Rather than chasing near-perfect fidelity, participants offer a further role for SBA as complementary to WBA. It is acknowledged that there are many procedure-specific competencies that cannot be reliably assessed in the workplace, either because they are not encountered with sufficient frequency, or to use their presentation as opportunistically to assess trainees would be a patient safety concern: *"in Ireland it's… like it's so ad hoc, as you know, it's so random, like you might be in [one training center] and get a car crash, someone coming in or you may not see one the whole way through your training"*. Trainers report a related lack of clinical time for assessment in many procedures: *"given increased trainee intake and lack of clinical time for assessment in the operating room, sim-based assessment is the way forward" [Trainer].* Ancillary benefits of standardizing training outcomes were anticipated: *"Simulation is a more reliable way of making sure that key factors in the curriculum are covered. Or you can make sure that people, who maybe were unfortunate not to come across more rare presentations of thing, at least they are not coming toward the end of their training having never had exposure to something" – Trainee.*

Participants extended the idea of SBA being used prior to in-theatre operating by junior trainees to using such an approach to assess more senior trainees prior to undertaking more complex procedures at a later stage: *"Like there are certain procedures that you'll only learn at a fellow level or beyond, you know, such as like maybe trans-anal TME or something like that in colorectal, in which case starting on the simulator model is totally appropriate and you know has the same benefit at that stage as it would have for an earlier training." – Trainee.*

Beyond making up for an assessment gap, simulating less frequently encountered procedures may be a better assessment of higher-order levels of competence, in particular the ability of trainees to adapt their accumulated skills to new procedures

or presentations: *“maybe using something […] unusual or rare, by putting trainees under pressure is a good way to assess you know, core skills and decision making under pressure.” – Trainee.* This is related to the concept that less familiar procedures require higher cognitive load to perform *“Because, yeah, you're as uncomfortable as the junior training is for what you consider a core procedure”.* The intended outcome of these assessments should be to ensure that trainees *“at least […] know the steps”,* and ideally as part of a train-to-proficiency model to prepare trainees for consultant practice: *“one of the comments from the faculty that day was that if you're going to assess people on that, you, you should be training people on that as well. And I would completely support that, you know, for major trauma situations or situations that we don't see often but we do need as general surgeons, for example, we do need to have exposure to it and know how to deal with it whenever we have to.”- Trainer.* SBA is therefore perceived as having a different, and complimentary, role to assessment in the workplace.

### 5.4.4 Discussion

Simulation-based assessment is perceived, by participants undertaking a novel simulation-based assessment of operative competence, as having a primary role in determining whether trainees have sufficient canonical knowledge and skill to undertake a supervised role as lead surgeon in the operating theatre. Difficulties in assessing contextual or personalized competence using simulation relate to the authenticity gap between simulation- and workplace-based assessments, driven by insufficient physical and psychological fidelity of experienced simulations. The lack of professional authenticity leads to an inability to capture a significant proportion of finer technical and non-technical skills deemed to be essential to determine competence in senior trainees. A superior ability of workplace-based assessment in capturing contextual and personalized competence is recognized, due to its obvious authenticity, and its longitudinal nature and reliability when repeated across a wide variety of operative contexts. Participants in this study put forward two proposals regarding the role of SBA in operative competence assessments: 1) Increasing the professional authenticity of SBA to approach real-world levels of fidelity, therefore

capturing a broader range of competencies, and 2) accepting the limitations of SBA in assessing higher levels of competence, instead using SBA to assess canonical competence in procedures encountered rarely in the workplace. The latter supports the use of SBA and WBA in a complimentary fashion.

Ten Cate et al. describe medical competence as a multi-layered construct with a distinct hierarchy[572]. The layers of this concept apply readily to learning a surgical procedure, with knowledge of key anatomy and operative steps required before applying such skills in the workplace, and further developing a personal style or approach[572]. Participants in this study concur with the implications of Ten Cate's conceptualization when it comes to assessment; contextual competence requires workplace observation, by multiple experts, in a variety of settings[572]. This further relates to established assessment theory in line with Miller's pyramid[576], where assessments at the 'shows how' level are readily conducted with simulation while assessing what a trainee 'does' requires direct observation in the workplace. Andreatta and Grupen, conceptualizing validity evidence for simulation, describe the hypothetical ideal assessment whereby the construct of interest (e.g. operative competence) is represented perfectly by the designed simulation[577]. Participants in this study perceive a construct-underrepresentation in experienced SBA, where designed stations, in this novel sim-based assessment at least, failed to capture the broad construct of 'competence'[577].

The findings of this study have implications for the use of simulation as an assessment modality in surgical training. Focusing on single-point, high stakes assessments of operative competence at certification level[14 15 83] risks assessment redundancy and construct under-representation. Furthermore, it fails to realize the potential of SBA in ensuring competence in less frequently encountered procedures as part of a formative, train-to-proficiency paradigm, ensuring a baseline level of competence in procedures that cannot be assessed in the workplace. Such procedures can now be readily identified by collating longitudinal trainee performance data[452 578].

*Limitations*

Assessment is context dependent, and the participants in this study took part in a novel simulation-based assessment of operative competence, designed to assess the performance of senior trainees in General Surgery. The experience of participating in this assessment, in a well-resourced and centralised simulation center, informed the perceptions experiences expressed in this study. While the findings may not be generalizable to other assessment contexts, we believe the experiences of these trainees and trainers can aid the development of future sim-based assessment frameworks in surgery.

**Conclusion**

Simulation-based assessment should be used in a complimentary fashion to workplace-based assessments. By recognizing its imperfect professional authenticity and the associated construct-underrepresentation with respect to operative competence, we can better design simulation-based assessment curricula to the benefit of trainers, trainees and their patients.

## 5.5 Additional Material: Why do residents fail simulation-based assessments of operative competence? A qualitative analysis

Reference: Toale, C., Morris, M. & Kavanagh, D.O. Why do residents fail simulation-based assessments of operative competence? A qualitative analysis. *Global Surg Educ* **2**, 83 (2023). https://doi.org/10.1007/s44186-023-00161-1

## Chapter 6: General Discussion

This chapter reviews the main findings of this thesis, and further integrates these findings to provide a blueprint for the use of technical performance assessments across surgical training, and in particular to inform certification decisions at end of training.

Role of thesis author: Mr. Toale authored this chapter.

## 6.1 Main Findings

### 6.1.1. Determining the need for simulation-based technical skill assessment in surgical training, in the context of a multi-modality programmatic approach to procedural assessment

There is a moral and ethical argument to be made regarding the need for simulation-based training and assessment in surgery. Simulation can be used to reconcile the competing interests of individual patients and broader society in training surgeons, when incorporated into training as part of a 'train-to-proficiency' paradigm. There are significant threats to the validity of currently utilised operative assessment modalities in the context of surgical training in Ireland and the United Kingdom (namely Workplace-Based Assessments and operative experience targets). Workplace-Based Assessments are used opportunistically by trainers and trainees, and are seen as a 'tick-box' exercise. There is limited evidence to support the use of operative experience targets as a surrogate measure of operative competence.

There is broadly positive agreement amongst surgical trainees and trainers, in the international published literature, regarding the use of simulation as an assessment modality in surgery, though its use in 'high-stakes' settings remains controversial. In our series of semi-structured interviews conducted with stakeholder representatives, it emerged that simulation could be acceptably used to ensure a 'minimum standard' of technical proficiency during early training, with a view to identifying under-performing trainees, and subsequently providing targeted remediation. It was important to interviewees that simulation-based assessment (SBA) should not replace the judgement of expert, trained, surgical trainers in high-stakes scenarios. These findings are further emphasised by trainee respondents to a national survey; simulation was perceived to be an appropriate modality when assessing a trainee's procedural knowledge, but deemed to be less useful when assessing 'higher-order' skills such as communication, teamwork, or decision-making. Further barriers toward the implementation of acceptable SBA were raised: lack of access to simulation and perceived deficiencies in the appropriate fidelity of experienced simulators.

### 6.1.2 Validity evidence for the use of objective technical aptitude assessments at the time of surgical trainee selection

The use of three objective assessments of technical aptitude (paper-based tests, pictorial surface orientation testing, and manual dexterity testing) were explored with a view to determining the relationship between derived scores and future performance. Aggregated aptitude or 'fundamental ability' scores were positively, and significantly, associated with future performance across simulation-based and workplace-based assessments of operative skill. These findings were independent of logged operative experience and undergraduate academic performance.

A further analysis of the Royal College of Surgeons in Ireland (RCSI) trainee selection process, incorporating technical aptitude assessments, was conducted. Composite scores incorporating measurements across technical and non-technical domains were associated with likelihood of subsequent progression to Higher Specialist Training (HST). On multivariable linear regression analysis, technical aptitude assessments were again positively associated with future performance in workplace-based and simulation-based technical skill assessments (β 0.31, 95% CI 0.14 - 0.48 and β 0.57, 95% CI 0.33 - 0.810 respectively).

### 6.1.3 Exploring the utility of an implemented simulation-based assessment designed to inform decisions on progression from Core Surgical Training to Higher Specialist training

At a Core Surgical Training (CST) level, there is evidence supporting the idea that workplace-based assessments of operative skill are used on an opportunistic basis. Submitted assessments cluster around submission deadlines, and trainees do not submit more than the mandated minimum number of assessments. These findings are a threat to the validity of workplace-based assessments. Furthermore, we have identified that a larger number of assessments than those which are currently mandated is required in order to achieve sufficiently high dependability indices for use in high-stakes assessment. Four assessors observing ten cases each would be required to provide a reliable judgement of a trainee's level of operative competence across an undifferentiated case-mix. Such large numbers of assessments may not be feasible in the workplace given observed assessment submission patterns and known barriers.

We further provide validity evidence for the use of an implemented simulation-based assessment of operative skill during Core Surgical Training, the Objective Surgical Skill (OSS) examination. Observation across 20 procedural stations would be required to reliably determine a trainee's level of operative competence. Currently implemented SBA is therefore insufficiently reliable to assess operative competence to a level required for high-stakes decision making; these findings support the use of a multi-modal, programmatic approach to technical skill assessment. The reliability of a single simulation-based assessment observation is, however, greater than the reliability of a single observation in the workplace. If the construct of interest is the determination of competence in a series of core procedural tasks or skills, SBA may therefore represent a more reliable approach.

### 6.1.4 Development of a novel simulation-based assessment designed to evaluate operative competence in senior general surgery trainees

Using a modified Delphi consensus approach (and modified Copenhagen Academy for Medical Education and Simulation [CAMES] needs assessment formula) we determined a prioritised list of procedures for use in simulation-based assessments of General Surgery Higher Specialist Training (HST). We subsequently developed an eight-station Objective Structured Clinical Examination (OSCE), using a variety of bio-hybrid and virtual-reality models. The Virtual Reality platform LapSim™ (Surgical Science, Sweden) was chosen to simulate laparoscopic procedures due to the wealth of published validity evidence regarding its use in simulation-based surgical training and assessment. This novel assessment could discriminate between trainees in year 1-4 and years 5-8 of training, with acceptable internal consistency metrics. A pass-fail benchmark resulted in all 'senior' trainees, and two of six junior trainees, passing the assessment.

This novel SBA formed the basis of a qualitative 'case-study'. On content analysis of written feedback, 'the model as a confounder' had important and unwanted effects on the scoring decisions of trainers. On thematic analysis of semi-structured individual and group interviews, conducted with participants and assessors, it emerged that SBA was viewed as having a primary role in determining canonical knowledge and skill, most valued prior to undertaking supervised 'real-world' operating. Participants envisaged a role for SBA in more senior training if used to determine canonical competence in procedures encountered infrequently in practice.

## 6.2 Incorporating Simulation-Based Assessment of Operative Competence into the Certification of Trainees: A Proposed Blueprint

**Abstract:**

**Objective**

The aim of this article is to outline a proposal for evidence- and theory-based incorporation of simulation as an assessment modality into the certification process, in the context of post-graduate surgical training.

**Summary Background Data**

Competency-based education principles have been integrated into surgical training programs. Such outcome-based curricula require the certification of graduate capabilities, which in turn requires valid and reliable methods of competency assessment. Simulation has been explored as a high-stakes assessment modality, though its role in contemporary surgical training remains to be fully established.

**Methods**

This article proposes a framework for incorporating simulation as an assessment modality into surgical training curricula, with reference to contemporary operative performance assessment science and theory.

**Results**

It is challenging for simulation to achieve sufficient professional authenticity to assess operative competence with sufficient confidence required for certification. Similarly, it is likely not feasible for simulation to assess performance across all procedures, and all components of all procedures, in which competence might be mandated by modern curricula. Workplace-based assessment can be reliably used to assess performance in commonly performed procedures. Simulation has a role in the assessment of high acuity, low frequency procedures in surgery, which are not encountered with sufficient frequency in the workplace to permit reliable assessment.

## Conclusion

Simulation as an assessment modality should not be used as a single-end point, end of training examination, but rather should complement workplace-based evaluations as part of a programmatic approach to training progression and end-of-training certification.

### 6.2.1 Competence, Competency-Based Education, and Surgery

The certification of trainees requires a defined standard for 'competence', against which performance can be compared. Several authors have attempted to define this elusive concept[592-594], facing criticisms of reductionism[595], construct under-representation[596] and incompatibility with theories of expertise attainment[597]. In the context of the United Kingdom and Ireland, "the end of training is reached when supervisors agree that a trainee is performing at the level of a day-one consultant, successfully managing the unselected emergency take, clinics and ward-care, operating lists and multi-disciplinary working while demonstrating the generic professional behaviours required of all doctors"[598]. While pragmatic, reference to a 'day-one consultant' does not bring us closer to defining what level of expertise is required at this level. Furthermore, this definition is insufficient when it comes to conducting assessments at a case or procedural level. This gap between lofty definitions of competence and the 'operationalisation' of assessment has led to the development of programmatic assessment frameworks, with 'Procedure-Based Assessment' (PBA)[599] being the primary tool for evaluating the performance of procedures in the workplaces of the United Kingdom and Ireland, and 'Entrustable Professional Activity' (EPA) assessments playing a similar role in the US-context[600].

Ten Cate et al. have recently proposed a conceptualisation of medical competence as a multi-layered construct, incorporating three distinct levels; 1) 'Canonical knowledge and skill', 2) Context-dependent knowledge, skills and attitudes, and 3) Personalised competence including personal skills, interests, habits and convictions[601]. This concept represents a useful lens with which to view operative competence attainment and assessment; core anatomical and procedural knowledge is required prior to developing skills in the workplace context-such as adaptability to between-case variations, interacting professionally with the operative environment and displaying teamwork and communication expertise. Personalised technique or 'style', innovation and development of specific operative interests might be thought of as 'the art of surgery'. Where simulation sits as an assessment modality within this example of a modern concept of competence will be explored.

### 6.2.2 Operative assessment in surgery

The traditional apprenticeship model of surgical training has been replaced by CBE, reliant on a suite of repeated longitudinal assessments, generally performed by several trainers across the continuum of surgical training. In the United Kingdom and Ireland, competence in the performance of procedures is informed by workplace-based assessments, multi-source feedback, and subjective trainer judgement through a 'multi-consultant report'[598]. Milestone assessments[373] and EPAs[600] inform competency committee decisions for US residents. Such workplace-based assessments are essential to competency-based education programs as naturalistic evaluations of *authentic* clinical encounters, supporting frequent and longitudinal assessment in context[602-605]. Threats to their validity as a summative assessment tool relate to concerns regarding objectivity, standardisability, and assessment gaming[430].

Given their summative role, EPAs and PBAs should further strive to achieve high levels of reliability. Achieving a reliable estimate of a trainee's level of operative skill requires repeated observations[606]; these should be undertaken by a number of assessors across procedural contexts to account for inter-rater and inter-case variation. The number of observations required to achieve high levels of reliability can be estimated using generalizability theory with mathematical modelling, and other statistical techniques[578 607 608]. Findings from articles reporting these methods support the idea that repeated observations of performance are required to assess operative competence; in Williams et al.'s 2019 study, 17 observations were required to reliably assess performance in laparoscopic cholecystectomy[608]. It is more difficult still to assess performance across pooled procedure types without large numbers of observations[608 609].

### 6.2.3 Simulation, assessment and certification

Simulation-based assessments have been developed to inform certification processes; for example, the Fundamentals of Laparoscopic Surgery examination is a required benchmark for US residents in a number of surgical sub-specialties[72-76]. At end-of-training certification, there are some notable examples of simulation as a high-stakes assessment modality. Then European Board of Surgery Qualification in Vascular Surgery (EBSQVS) includes assessment of procedural skill[84]. De Montbrun and colleagues have described the development of the General Surgery Objective Structured Assessment of Technical Skill (GOSATS)[15] and the Colorectal Objective Structured Assessment of Technical Skill (COSATS)[14 18 86]. These assessments, designed as technical skill certification examinations, have undergone validity testing using contemporary frameworks with favourable results[14 15 18 86]. The implementation of the COSATS as a National Certification Examination was described in 2016[18].

Simulation as an assessment modality has a number of potential advantages over workplace-based assessment[610]. Patient safety concerns, particularly with respect to high acuity procedures, are avoided. Assessments are more readily standardisable. Assessment opportunism is negated. The use of SBA to inform certification decisions however, remains controversial[294]; stakeholders report concerns regarding the potential for SBA to overrule expert trainer judgement, and emphasise the value of SBA as a formative, rather than summative, assessment modality[329]. Trainers and trainees are aware of a deficit in authenticity when comparing sim-based to workplace-based performance assessment.

### 6.2.4 Professional authenticity and simulation-based assessment of operative competence

In spite of the proposed benefits of simulation as an assessment modality, significant barriers with respect to its implementation remain. Of primary concern is the difficulty in producing simulations of sufficient professional authenticity (or procedural fidelity) to assess operative *competence*. Competence, as defined above, is a broad construct incorporating diverse domains such as psychomotor skill, declarative knowledge, interpersonal skills, personal resourcefulness, and advanced cognitive skills[611]. The latter, relating to such skills and attributes as a surgeon's adaptability to complex or unfamiliar scenarios as well as their intra-operative judgement and decision making, are perceived by trainers and trainees as being fundamental when determining the competence of graduating trainees.

Andreatta and Grupen describe the validity of a given assessment in the context of what it is intended to measure, i.e. the construct of interest[577]. In the case of high stakes SBA, the construct which simulation seeks to represent is the competent performance of operative procedures. A perfect assessment would reliably capture and assess the broad range of knowledge, skills and attributes required to perform a given operation, or group of operations. Considering Ten Cate at al.'s multi-layered conceptualisation of medical competence, such assessments at certification level might be expected to assess not only canonical competence – procedural knowledge and specific skills, but also *contextual* competence, with emphasis on the application of such knowledge and skills within the opportunities and constraints of the theatre environment[601]. We can imagine some of the specific knowledge, skills and attitudes such assessments might encompass; the adaptation of surgical skills to challenging or unusual cases, the recognition and management of variant anatomy, the peri-operative management of surgical conditions in the context of a holistic view of patient health, and respectful and effective communication with the operating team. Such assessments are required to display high levels of professional authenticity, i.e. the models and simulation must display high levels of physical, functional, psychological and cognitive fidelity. The assessment of knowledge, skills and

attitudes in context requires the assessment to simulate that context, the operating room, to a sufficient degree that candidate's interactions with a simulated patient, case, team, and theatre environment can be reliably assumed to represent the behaviour of trainees in the workplace (figure 6.2.1). This concept relates to the traditional view of hierarchical assessment according to Miller's pyramid, and a recognition that simulation assesses at a level of 'shows how', rather than 'does'[576]. The ability of simulation to represent the operating environment, and therefore the construct of competence, must be considered beyond traditional quantitative measures of validity. For example, senior trainees would be expected to perform better than more junior trainees in simulation-based Objective Structured Clinical Examination (OSCE), and may approach or exceed expert benchmarks of operative skill. Acceptable internal consistency and reliability metrics could be reached, with a reliable pass-fail score generated. Without a simulation that approximates the real-world environment to a high degree however, such assessments risk capturing *skill* rather than *competence.* This has significant implications for the use of simulation as a high-stakes assessment modality.

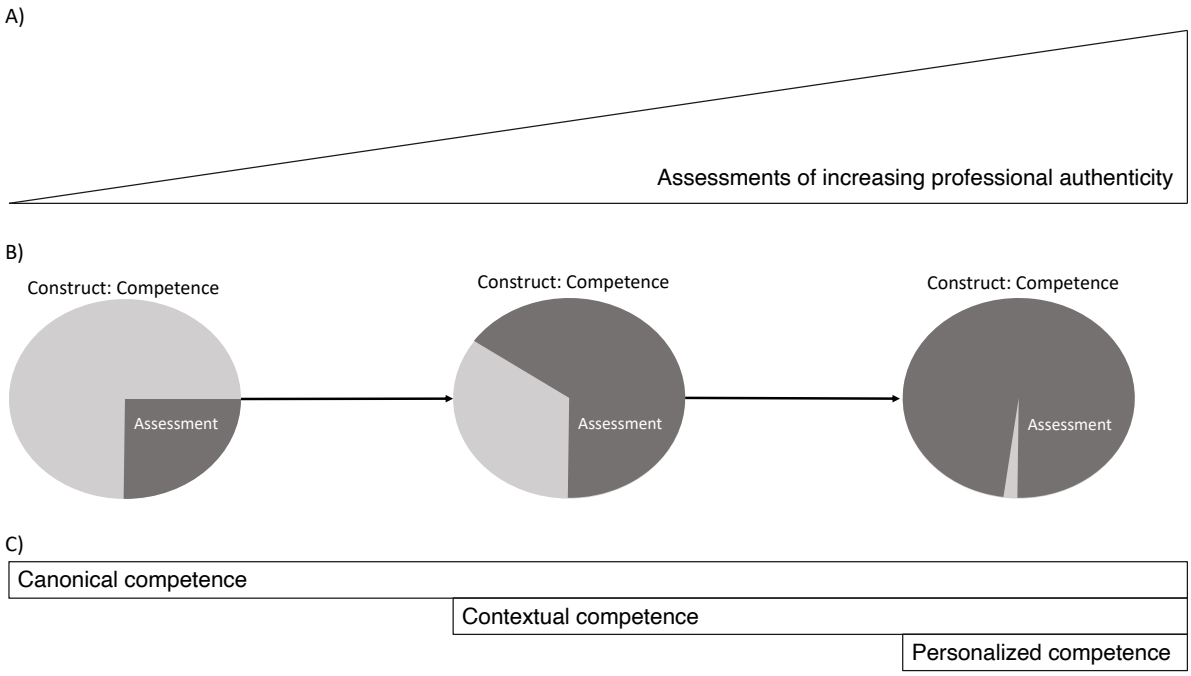



*Figure 6.2.1: Assessments of increasing professional authenticity (A) capture a broader range of knowledge, skills and attitudes required to represent the construct of 'competence' (B), in turn leading to increasing validity when assessing competence along the continuum from canonical competence through contextual competence to personalized competence (C). Adapted from: Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990 Sep;65(9 Suppl):S63-7. doi: 10.1097/00001888-199009000-00045. PMID: 2400509, Andreatta PB, Gruppen LD. Conceptualising and classifying validity evidence for simulation. Med Educ. 2009 Nov;43(11):1028-35. doi: 10.1111/j.1365-2923.2009.03454.x. PMID: 19874494, and Ten Cate O, Khursigara-Slattery N, Cruess RL, Hamstra SJ, Steinert Y, Sternszus R. Medical competence as a multi-layered construct. Med Educ. 2023 Jul 16. doi: 10.1111/medu.15162. Epub ahead of print. PMID: 37455291.*

### 6.2.5 The simulated and workplace environments; two views on 'construct under-representation'

We can hypothesise that there is an authenticity deficit that exists between assessments conducted in the simulated environment and assessments conducted in the workplace context, and that this authenticity gap can lead to construct underrepresentation. By increasing the fidelity of simulated operative cases, the authenticity of workplace based assessment can be reached. A simulation can be imagined that is of sufficient psychological, physical and cognitive fidelity that the nuances of naturalistic workplace behaviour can be replicated, observed and assessed to a degree of sufficient validity that competence in the workplace can be implied by performance in a simulation-based assessment. Such assessments in future may become more feasible due to improvements in simulator technology[612], the use of in-situ simulation[613], and team or interprofessional sim-based assessment[614]. A further challenge still, however, is the difficulty in using simulation to conduct repeated, longitudinal assessments, particularly in the current surgical training paradigm where training and service-provision are conducted simultaneously.

In this respect, workplace-based assessments have a clear advantage over simulation-based assessment, in permitting frequent and longitudinal assessment. As previously discussed, a repeated number of observations are required to reliably assess performance in the workplace. This number is possible to estimate using observational data and statistical modelling[578 607 608]. It is conceivable that trainees could be observed performing a sufficient number of commonly encountered procedures for pooled performance assessments to be considered reliably representative of their *competence* in performing said procedure[452]. Conversely, there are several procedures for which such an assessment model is not sufficient or feasible. For example, it is imaginable that there are several procedures which will not be encountered sufficiently frequently during the average trainee's residency to permit reliable procedure-specific competence assessment.

Therefore, even workplace based assessment may struggle to represent the construct of competence across all procedures in which proficient performance might be expected in independently practicing surgeons. It has been proposed by Williams et al. (2021) that performance in these low-frequency procedures should be assessed in aggregate[452]. This idea relies in part on the intuitive assumption that there are shared domains of knowledge and skills that transfer from low- to high-frequency procedures and vice versa. In this way, a certain degree of canonical and contextual competence may be implied in once procedure by observing a trainee's performance in other, more frequently performed operations[452].

### 6.2.6 High acuity, low occurrence procedures in surgery

A subset of these rarely performed procedures are high-acuity procedures that, if used for opportunistic assessment of independent trainee performance, would represent an unacceptable patient safety risk. Furthermore, competence in some of these procedures might not necessarily be implied by skills and knowledge obtained from the performance of more frequently encountered procedures. A general surgery trainee displaying competent performance in core, common procedures (appendicectomy, cholecystectomy, inguinal hernia repair etc.) may not necessarily possess sufficient anatomical knowledge, procedural knowledge or procedure-specific skill to safely and competently perform such procedures or manoeuvres as creating a surgical airway, performing a trauma thoracotomy, or cross-clamping the supra-coeliac aorta for control of catastrophic bleeding. The high-frequency and low-frequency procedures are sufficiently different that significant overlap in acquired knowledge and skills cannot be assumed.

The value of simulation in teaching for high acuity, low frequency scenarios has been explored in the health professions education literature[615-617]. Such procedures lend themselves to simulation, allowing for repeated deliberate-practice in a train-to-proficiency paradigm, with opportunity for expert feedback and formative assessment. Furthermore, it allows for training and assessment of such procedures to take place in a safe environment, with no risk to patients.

### 6.2.7 Simulation-based procedural assessment and its role in certification

Stakeholders support the use of simulation as an assessment modality when used in early surgical training, to support patient safety and safe trainee autonomy in the transition to in-theatre operating[329]. Current limitations, explored above, may limit its ability to reliably assess higher levels of competence required to certify more senior trainees, particularly when the authenticity of SBA is compared to assessment in the workplace. We propose, with reference to William's et al.'s 'a proposed blueprint for operative performance, training, assessment and certification', the following role for simulation-based assessment in the certification of procedure-specific competencies in surgical training.:

- 1) Competence in commonly performed procedures should be assessed in the workplace, provided that these procedures are encountered with sufficient frequency to permit reliable assessment in accordance with statistical principles of assessment reliability.
- 2) Competence in all procedures should be assessed, at least formatively. Performance in 'Mid- frequency' procedures, i.e. procedures likely to be encountered during training, but not at a sufficient frequency to permit reliable assessment along traditional conceptualisation of reliability, can be assessed on aggregate, under the assumption that assessment still contributed to perceptions of a candidate's operative competence across shared domains of knowledge, skills and attitudes across procedures.
- 3) Lower frequency procedures can be defined as procedures deemed essential for surgical training that are not encountered sufficiently frequently to permit reliable and/or safe assessment in the workplace. A prioritised list of such high-acuity, low-occurrence procedures should be identified and assessed using simulation. This assessment should take place as part of a simulation-based train to proficiency model, with a view to improving patient safety and ensuring standardisation of graduate outcomes across unique training experiences and contexts. Conceptually, such a program would train and assess procedural competencies to at least a level of *canonical competence*, with candidates displaying a proficient level of procedure-

specific knowledge and skill. High fidelity simulation may even permit the assessment of such skills to a level approaching *contextual* competence. It may be assumed that a substantial knowledge of transferrable, non-procedure specific knowledge, skills and attitudes will have been ensured through assessment of competence in frequently encountered procedures, to mitigate construct-under representation of the assessment framework as a whole.

### 6.2.8 Conclusion

Simulation-based assessment may struggle to feasibly replicate the operating environment to a sufficient level of professional authenticity to assess operative competence, particularly beyond the level of canonical competence. The assessment of higher levels of competence, i.e. contextual competence, is required for the certification of graduating surgical trainees. Assessment at this level is more authentically conducted in the workplace, with frequent, repeated assessments in a variety of case- specific contexts mitigating against potential construct under-representation. The reliable assessment of competence at this level requires a substantial number of procedure-specific observations; this is feasible in the workplace for a small number of high-frequency procedures. Competence in one procedure cannot necessarily be assumed by the certification of competence in another as skill transferability is not guaranteed, particularly when said procedures differ substantially in their knowledge or skill domains. Workplace-based assessment therefore may fail to ensure competence in high-acuity, low occurrence procedures. These procedures can be trained and assessed in a simulated environment to ensure a degree of proficiency in graduating trainees who may be reasonably expected to perform these procedures as an independent practitioner. SBA should therefore form part of a programmatic approach to operative competence assessment, throughout training and at certification level.

## 6.3 High-stakes technical performance assessments across the continuum of surgical training

**Abstract**

Objective:

The aim of this article is to outline a proposal for the use of high stakes technical performance assessments, defined as assessments of technical skill that take place outside of the operating theatre, across the continuum of surgical training.

Summary background data:

Technological advancements have allowed for the assessment of operative technical skill outside of the workplace environment, developed and evaluated in line with modern concepts of assessment validity. With the introduction of competency-based education principles across surgical training curricula, increasing scrutiny is being placed on the validity and reliability of assessments used to inform high-stakes training decisions regarding selection, progression through training, autonomy granting and end-of-training certification. The role of technical performance assessments, conducted outside of the workplace environment, is yet to be fully established.

Methods:

This article discusses the use of high-stakes technical performance assessments across the continuum of surgical training, with reference to relevant theories of surgical skill acquisition and assessment, along with contemporary evidence from the performance science literature.

Results:

Assessments of increasing professional authenticity are required across the continuum of surgical training. In an attempt to capture inherent ability rather than skill at the time of trainee selection, assessments should be of deliberately low levels

of professional authenticity. Assessment in early surgical training should focus on ensuring the acquisition of core surgical skills, and then assessing for canonical competence across a series of core, commonly encountered procedures. Baseline competence in these procedures should ideally be assessed prior to the transition to (supervised) lead operator, with a view to ensuring patient safety and improving the efficiency of in-theatre training. Assessment later in training relies more heavily on the workplace, with simulation offering a complimentary role in ensuring a minimum competency standard in procedures less likely to be encountered in the workplace.

Conclusion:

High-stakes technical performance assessments can play an important role as part of a programmatic approach to surgical training and certification. They can and should be designed with an appropriate level of professional authenticity to ensure adequate and appropriate representation of the assessment construct.

### 6.3.1 Surgical Skill and the Acquisition of Operative Competence

Surgery is somewhat unique across the medical subspecialties, in that certification of competence requires a greater focus on the acquisition and assessment of technical skill. Skill development can be described using a number of theories from the social sciences; Fitts and Posner's 3-stage theory of motor skill acquisition puts forward that the mastery of a technical skill requires the learner to progress through three distinct stages: 1) the cognitive or learning stage, 2) the associative or motor behaviour stage, and 3) the autonomous or expert stage[618]. Competence is achieved through training and deliberate practice, with skill observation, assessment, and expert feedback[619 620]. It is in turn influenced by learner behaviours such as the degree to which they engage with repeated deliberate practice[286] or mental rehearsal[621]. Traditionally, surgery has relied on a time-based educational structure, informed by an apprenticeship model of training and assessment[622]. Increasingly, competency-based education curricula have been introduced across surgical training jurisdictions[430 431 623], with a focus on time-independent certification of graduate outcome capabilities, in line with principles of mastery learning. In this article we discuss the use of technical performance assessments across surgical training in line with competency-based education principles. We define technical performance as assessments of technical aptitude, skill, or competence that take place outside of the operating theatre context. These include simulation-based assessments, but also assessments of technical aptitude that do not attempt to simulate real-world surgical tasks or procedures.

### 6.3.2 Mastery Learning, Competence and Surgical Training

Mastery learning concepts are congruent with the principles of competency-based education. In essence, mastery learning requires that trainees are taught and assessed against rigorous knowledge, skill and affective attributes in a time-independent fashion, with deliberate practice and remediation undertaken by those who fail to achieve such standards until they are met[624]. Seven key features of mastery learning curricula are described[624 625]:

1) baseline assessment
2) learning objectives sequenced in order of increasing difficulty
3) engagement in educational activities such as deliberate practice
4) a pre-defined minimum passing standard
5) formative assessment to gauge achievement of an educational outcome, followed by
6) advancement to the next educational outcome once the pre-defined standard is reached, and
7) continued practice until the pre-defined mastery standard is reached[625].

Central to mastery learning is the definition of a minimum passing standard. In competency-based curricula, this requires a definition of 'competence' ; in surgery, it further requires a working definition of what competence means in the context of technical, or more broadly 'operative' skill. Defining competence has proven to be elusive. Comprehensive definitions such as 'the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served' have been proposed[592]. While capturing a broad range of knowledge, skill and attribute domains expected of independently practicing surgeons, such definitions are difficult to operationalise when it comes to the assessment of a given trainee when performing a particular procedure or skill. This has led to the development of high level outcomes such as 'generic professional capabilities' and 'capabilities in practice' in the United Kingdom context[374] and 'milestones'[373] in the United States. Performance in individual procedures or tasks is assessed against these pre-defined outcomes using tools such as 'Procedure Based Assessments (PBAs) in the United Kingdom and Ireland[599], and 'Entrustable Professional Activity' (EPA) assessments in the US context[600].

Ten Cate et al. (2023) have recently developed a conceptualisation of competence as a multi-layered construct incorporating three distinct layers, acquired in a hierarchical fashion across medical training[601]. The first layer is 'canonical competence' – context-independent knowledge and skill. This is followed by

'contextual competence' with an emphasis on the application of these skills in the clinical workplace environment. Finally, a third layer of 'personalised competence' is outlined, focussing on the development of a personal style of health care provision, grounded in adequate knowledge and skill[601]. Such a framework is readily applicable to surgical training and has implications for assessment across the continuum from selection to certification. It is postulated that canonical competence is readily assessed using standardised methods, such as simulation, but that contextual competence, as the name suggests, is better assessed in the workplace environment.

### 6.3.3 Assessment across the Continuum of Surgical Education

Though postgraduate surgical curricula differ across jurisdictions and sub-specialties, there are certain core and common features related to skill assessment. The first is the requirement for some form of selection or recruitment of surgical trainees, at which point technical skill or aptitude is inconsistently assessed[378]. The next is the process of progression through training, with technical skill either formally or informally assessed through a variety of modalities and tools. Finally, a judgement is required to be made regarding the ability of a given trainee to independently operate at the point of certification. The Irish postgraduate surgical training programme, administered by the Royal College of Surgeons in Ireland, can provide an example (Figure 6.3.1).

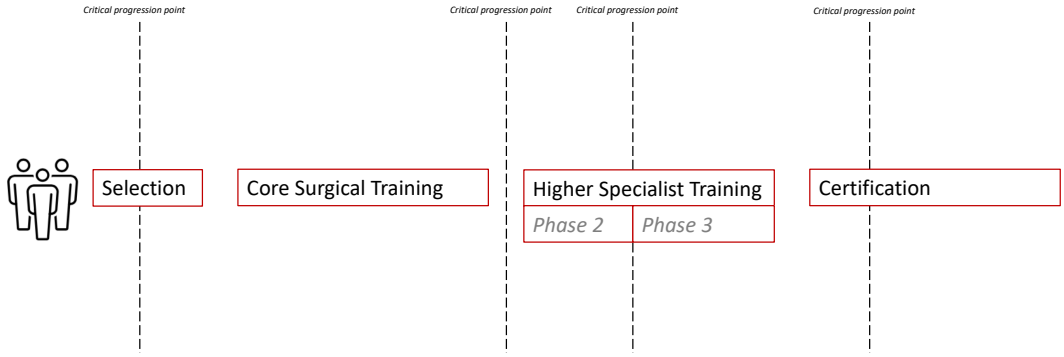



*Figure 6.3.1: A simplified representation of the Irish surgical training curricula. The General Surgery curriculum 2021 outlines two distinct phases of Higher Specialist Training; at the end of phase one (indicative time four years) trainees are expected to have achieved competence 'to the level of a day-one consultant' in a variety of core emergency and elective general surgery procedures, before developing further competence in an area of sub-specialty interest during phase 3.*

Prospective candidates undergo a centralised, standardised multi-faceted selection process incorporating assessments of technical aptitude, scored performance in undergraduate education, and performance in a multi-station mini-interview. Successful candidates then undergo two years of undifferentiated 'Core Surgical Training', rotating through various clinical posts, with technical skill assessed using a combination of simulation- and workplace-based assessments. Candidates then apply for 'Higher Specialist Training' in their chosen field, nominally a six year program during which operative competence is primarily assessed in the workplace. On successful completion of Higher Specialist Training, informed by performance in a variety of technical and non-technical assessments in conjunction with performance in the Intercollegiate Fellowship Examinations, trainees are awarded a Certificate of Completion of Specialist Training (CCST). We will now discuss how assessment outside the operating theatre can be incorporated into such an assessment programme with reference to modern conceptualisations of competence, assessment validity and mastery learning principles.

### 6.3.4 Assessing technical skill outside of the operating room: representing the appropriate construct

Essential to the development of any technical performance assessment is the definition of the construct, and its interpretation[626]. In the context of end-of-training certification in general surgery, for example, we might expect 'competence' to be the construct of interest, defined as the level expected of a 'day-1 consultant' who can 'managed patients presenting with the full range of emergency General Surgery conditions and elective conditions in the generality of General Surgery'[374]. With reference to ten Cate et al,'s conceptualisation of competence, this would require certification of the *contextual* competence of trainees. As we have previously discussed, such assessments would need to capture a broad range of knowledge, skills and attitudes across a variety of applied clinical contexts. Performance assessments, unless they replicate the operating room environment and fidelity of tissue to a sufficient level of professional authenticity, risk under-representing this construct[577 627]. Assessing procedural competence requires an acknowledgment that domains beyond technical ability – teamwork, communication, and judgement for example – must be acknowledged, observed, and 'measured'. Even high-fidelity simulations are unlikely to capture the broad range of clinical scenarios encountered in the workplace, making workplace-based assessment potentially more meaningful at senior levels of training[601]. Workplace-based assessments further have the advantage of being more feasibly conducted in a longitudinal, repeated fashion.

It is important also to recognise the reverse of this concept; over-representation of the intended construct can lead to less meaningful, and less valid assessment[627 628]. Assessments should not be so complex that they are beyond the expected level of a trainee's learning[627]. Not only do overly-complex assessments unnecessarily assess skill that has not yet been learned, they also increase cognitive load leading to performance in the intended task to suffer[627]. Take, for example, technical aptitude assessments used to inform selection onto surgical training. In RCSI's selection framework, such aptitude assessments make no attempt to replicate real-world procedures or tasks[629]. Such assessments are not intended to assess competence,

or even ‘skill’, but rather aptitude or learning potential[630 631]. A high-fidelity, full procedural simulation-based assessment in this context would be meaningless, as candidates would not and should into be expected to have sufficient learned skill. Even simple simulations of technical skill such as suturing, for example, could qualify as construct over-representation in this context, as assessments would potentially favour candidates with increased operative *exposure* rather than *ability*. In early training, assessments of such core skills as suturing are entirely appropriate as candidates would be expected to have acquired these skills through their training experience. Again, full procedural simulation in the context of early or ‘Core’ training would risk increased cognitive load and unnecessarily complex assessments of procedural planning, intra-operative communication and advanced manoeuvres[577]. Increasing cognitive load could in turn affect performance in the intended construct, e.g. suturing[577]. Furthermore, increasing the professional authenticity of performance assessments invariably leads to feasibility concerns and greater costs[632], further emphasising the need for simulations of appropriate fidelity at each stage of surgical training[633].

### 6.3.5 Technical Performance assessments across the continuum of surgical training

In light of the above considerations with respect to mastery learning, competency-based education principles and assessment validity, it follows that the professional authenticity of technical performance assessments needs to increase along the continuum of surgical training from selection to certification, in order to appropriately capture the intended construct(s) at each stage and minimise the risk of construct under- or over-representation (Figure 6.3.2).

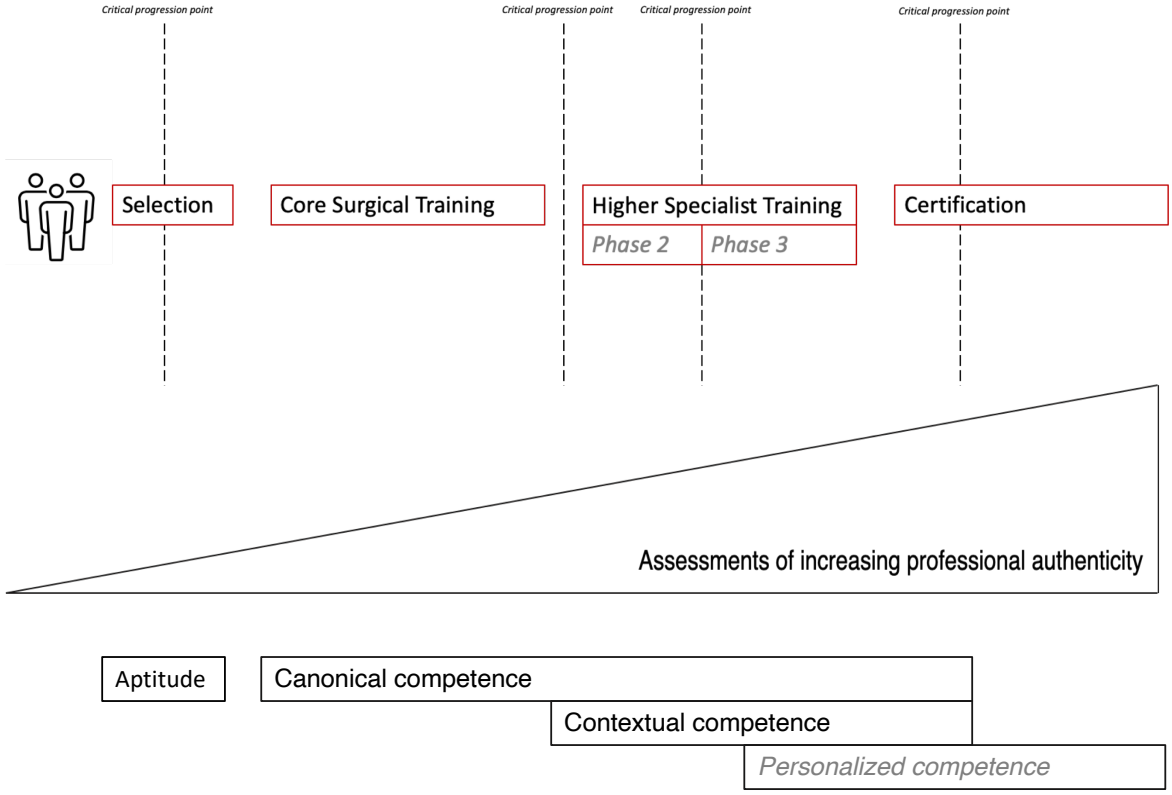



*Figure 6.3.2: The professional authenticity of assessments should increase along the continuum of surgical training from selection to end-of-training certification*

Given current technological, cost and feasibility constraints related to simulation-based assessment of operative competence, it is unlikely at present that simulation can be used to assess the broad range of context-dependent knowledge, skills, and attitudes required for certification as an independent performer of operative procedures at certification level. While such standardised assessments are valuable in determining canonical competence, workplace-based assessment, informed by the judgement of expert assessors, may provide a more meaningful assessment of contextual competence at a senior level of training (Figure 6.3.3).

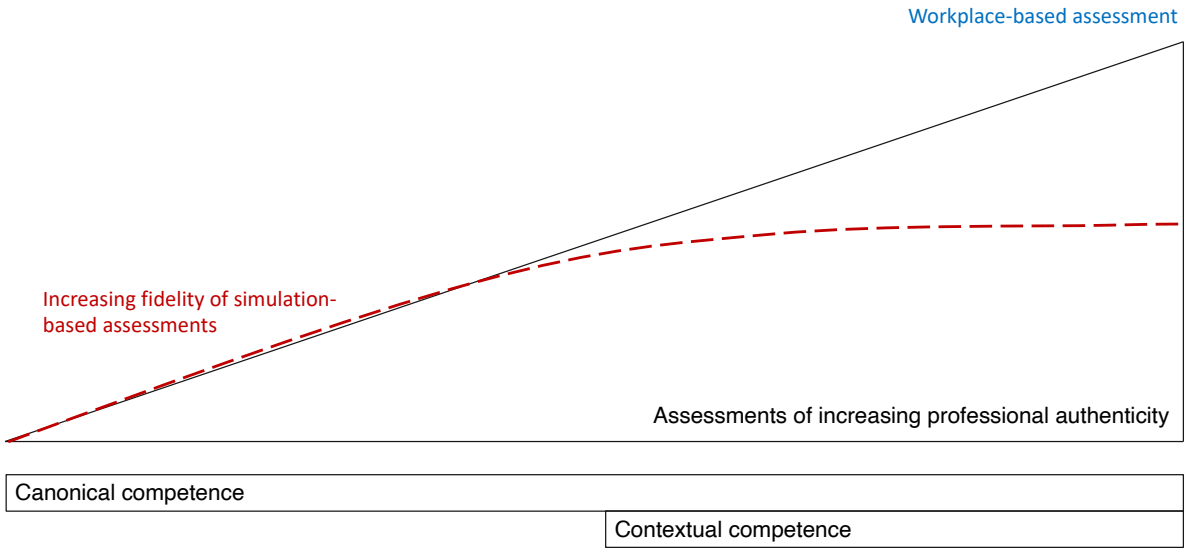



*Figure 6.3.3: Current simulation technology limits the ability to assess contextual competence due to deficiencies in fidelity and therefore professional authenticity*

### 6.3.6 Proposals regarding the use of technical performance assessments in surgical training

#### *6.3.6.1 Selection and Recruitment*

At selection, technical performance assessments should purposefully be designed to be of low professional authenticity, and should not attempt to replicate real-world procedures. This should be implemented to avoid construct under-representation, with the intent of assessing *aptitude* rather than skill, competence or related higher constructs. Aptitude should be interpreted as displaying sufficient innate technical ability such that competence in operative procedures is likely to be achieved during surgical training, over a reasonable timescale. Performance in such aptitude assessments should ideally *not* improve with increasing operative exposure. As technical aptitude is just one desirable attribute of many when selecting surgical trainees[634], such assessments should form part of a multi-factorial approach to recruitment where utilised, and should demonstrate predictive validity[635]. As a core principle of mastery learning is that all learners should be capable of becoming

proficient in a given skill given sufficient repeated deliberate practice, we do not propose that a 'cut-off' score derived from such assessments is used for categorical selection decisions.

*6.3.6.2 Progression through training*

In early surgical training, technical performance assessments should be of appropriate fidelity to assess the intended construct. Initially, such assessments will likely focus on ensuring sufficient level of competence in core surgical skills such as suturing, knot tying, basic laparoscopic skills and others. Increasing simulator fidelity can then lead to the assessment of *canonical competence* in full procedures or more complex skills. Simulation-based assessments at this level have an advantage over workplace-based assessment in this context as they are more readily standardisable. This in turn allows training programs to ensure that junior trainees, with heterogenous training experiences, have acquired the necessary technical skills and procedural knowledge to progress to more senior levels of training. Given the substantial published validity evidence regarding train-to-proficiency models in surgical training[100 636 637], we have previously argued that such training and assessment take place, where possible, prior to undertaking a supervised role as lead operator in the theatre setting[638]. At this point in training, the focus should be on assessing competence in skills and procedures that are frequently encountered in the workplace context. This has proposed benefits for patient safety, though such outcomes are difficult to ascertain[639]. Further hypothesised positive impacts relate to the efficiency of surgical training; ensuring canonical competence prior to in-theatre operating could allow theatre training time to be used for teaching and assessing context-dependent knowledge, skills and attitudes. Whether performance in such assessments could impact upon autonomy granting decisions or entrustment remains to be further studied.

In RCSI, simulation-based assessments of core, undifferentiated operative skills inform decisions on competitive progression from Core Surgical Training to Higher Specialist Training in the candidate's surgical specialty of interest. In recognition

once more that technical performance represents just one domain of competence desirable in trainees at this critical progression point, progression to Higher Specialist Training is decided on the basis of a programmatic approach to assessment of technical skills, non-technical skills and professional development[43].

*6.3.6.3 Certification*

We have previously argued that workplace-based assessment represents a more meaningful tool for assessing contextual competence in senior levels of training. This is in part due to the difficulty in feasibly conducting repeated longitudinal assessments using simulation. Furthermore, there is a perceived inability of current simulator technology to replicate the complexities and nuances of the operating environment to a sufficient degree to reliably assess contextual competence. This does not mean, however, that performance assessments have no role at this level. Rather, instead of focussing on single time point high stakes technical skill examinations[14 15], we propose that simulation-based assessment at this level should focus on the certification of canonical competence in procedures or skills in which a degree of proficiency is expected but which are encountered insufficiently frequently, or are of too high-acuity, to permit safe and/ or reliable assessment in the workplace.

**6.3.7 Looking to the future**

The hierarchical assessment concept of baseline aptitude testing, simulated skill assessment, and simulated full procedural assessment prior to supervised in-theatre operating may be familiar to those who have undertaken robotic surgery training[640]. Industry led training curricula, by coincidence or design, follow similar principles to those outlined above, likely informed by theories of mastery learning[641]. It is interesting that pre-procedural certification of a degree of competence prior to performing procedures for patients is accepted, and indeed mandated, in robotic surgery. This framework has yet to be applied across other surgical modalities. This is perhaps due to the current training structure which requires surgical trainees to provide a patient service in conjunction with their training in open and laparoscopic

surgery. In any case, the mastery learning, assessment validity and competency-based education principles herein provide theory- and evidence-based support for these curricula, and also provide for a curricular framework by which future robotics training, for example, could be delivered across the continuum of surgical training (Figure 6.3.4).

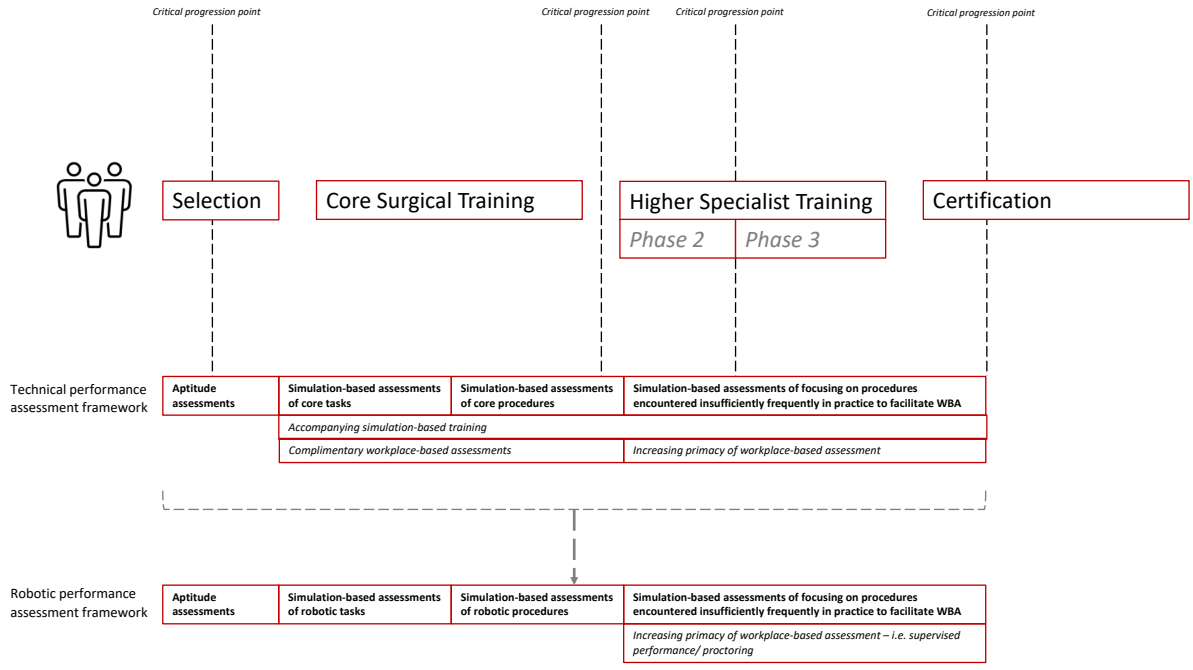



*Figure 6.3.4: Framework for the use of technical performance assessments across surgical training, with reference to a proposed assessment framework for robotic surgical training across the continuum from selection to certification*

### 6.3.8 Conclusions

Competency-based education in surgical training requires reliable and valid assessment of operative competence. Modern conceptualisations of competence frame it as a complex and multi-layered construct, with ten Cate et al. (2023) describing three distinct hierarchical layers – canonical competence, contextual competence, and personalised competence. Before competence is even considered, the construct of technical aptitude or ability can be considered and assessed using technological assessments. Assessing higher levels of operative competence requires assessments of increasing professional authenticity, while conversely

constructing assessments that are unnecessarily complex risks construct over-representation which in itself is a threat to assessment validity. We have therefore outlined proposals for the appropriate use of technical performance assessments as part of a programmatic approach to selection, progression through training and end-of-training certification, with reference to contemporary conceptualisations of mastery learning, assessment validity and competency-based education.

## 6.4 Future Directions

The findings from this thesis give rise to several questions which could form the basis of future research.

### 6.4.1 Using technical aptitude assessments to inform selection decisions

We have demonstrated that measures of technical aptitude are positively associated with the future performance of surgical trainees in early surgical training, and have further demonstrated how these assessments can be used to inform selection as part of a multi-modality approach to trainee recruitment. Whether these positive associations extend beyond the early years of surgical training should be studied. Furthermore, these aptitude assessments rely on the assumption that they measure innate aptitude as opposed to accrued fine motor skill or dexterity. For this reason scores achieved in these assessments should not improve over the course of surgical training; this assumption can and should be tested. Furthermore, the reliability of these assessments (i.e. how many observations or iterations are required to make a summary judgement of a trainee's innate ability) should be further explored by future studies. Finally, the acceptability of such an approach to trainee selection should be explored in line with principles of Social Validity.

### 6.4.2. Using SBA to inform progression through training

The validity of RCSI's assessment curriculum during Core Surgical Training has been previously explored. Given the ethical, theoretical and evidence-based recommendations outlined in this thesis, should trainees be mandated to reach a degree of proficiency in key procedural tasks or skills prior to undertaking supervised 'in-theatre' operating as the lead surgeon? What would the acceptability and validity of such an approach be and would there be subsequent measurable impacts on patient safety, trainee autonomy or trainee learning? Can we use simulation-based assessment to ensure that key training outcomes are met by all Core Surgical

Trainees, regardless of their unique and context-dependent operative experience across rotations?

### 6.4.3. Using SBA to inform end-of-training certification decisions

We have demonstrated that trainees and trainers are skeptical regarding the value of a single time-point, high-stakes simulation-based technical skill assessment being used to inform certification decisions. Future studies can explore whether a technical skills assessment deployed earlier in Higher Specialist Training, modelled on the assessment framework outlined in this thesis, can be used to meaningfully identify trainees in need of remediation early in their specialty training. Furthermore, it remains to be seen whether adopting a mastery-learning approach to such assessments, with trainees allowed to repeat until a certain standard is met, would lead to better training or patient-care outcomes.

Future simulator technology will determine whether sufficient fidelity can be reached to approach the authenticity of workplace-based assessment. Future studies could incorporate in-situ assessments with a team-based approach in order to improve the psychological fidelity of SBA. The utility of such assessments could be subsequently evaluated in order to determine whether they could be used to inform decisions regarding a trainee's level of contextual competence. Finally, trainers and trainees have suggested a role for SBA as complementary to workplace-based assessments. SBA could be used to assess competence in procedures which are encountered too infrequently, or are of too high acuity, for assessment to be reliably and/or safely conducted in the workplace. Delphi methodology and longitudinal performance data could be used to identify these procedures. A simulation-based training and assessment curriculum could be devised with a view to ensuring the competence of graduating trainees in a series of high-acuity, low-frequency procedures. Whether such an approach would improve the validity of a programmatic approach to operative competence assessment remains to be studied. Determining the benefits of such an approach on patient and educational outcomes represents a daunting challenge, but should nevertheless be considered by future studies.

## 6.5 Methodological Considerations

The methodological strengths and limitations of this thesis will now be further considered.

### *6.5.1 Strengths:*

Presented reviews were conducted and reported in line with best practice as outlined by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRSIMA) guidelines[642] and PRISMA- Scoping Review (ScR) extension[643]. Exploring the perspectives of key stakeholders and trainees toward simulation-based assessment followed an iterative approach as outlined by Artino et al. in "Developing questionnaires for educational research: AMEE Guide No. 87"[644]. Perspectives and experiences on simulation-based assessments were sought from a broad range of stakeholder representatives, including a patient representative. Findings from qualitative research are reported in line with established COnsolidated criteria for REporting Qualitative Research (COREQ) guidelines[305].

The studies outlined in 'Chapter 3: Validity evidence for the use of objective technical aptitude assessments at the time of surgical trainee selection' are sufficiently powered, and are reported in line with recommendations from the STrengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement 2007[349]. Potential predictor variables were recorded prior to outcome variables, and attempts were made to statistically account for recognized confounders.

In Chapter 4, we use longitudinal trainee assessment data from an established Core Surgical Training program to determine use patterns for workplace-based assessments of operative competence. A recognized 'Task, Gap, Action' framework is used to formally evaluate the feedback quality of formative assessments. The large number of submitted assessments by a large number of trainees and trained assessors is a strength in this regard. The use of generalizability theory to explore the reliability of assessments in this chapter deserves further exploration. The

strength of generalizability theory is that, unlike in 'classical' reliability theory, sources of measurement error can be investigated by one single study or experiment. In the absence of accepted 'sample size' guidelines for such studies, we have restricted the exploration of procedure-specific assessment reliability to procedures which have been completed by more than 50 trainees, in line with suggested practice from published literature[440]. As the function of workplace-based assessments of operative competence in a competency-based training program should be to determine the competence of a given trainee as an absolute, and not in relation to their colleagues, the index of dependability Φ was utilized instead of a 'G' statistic when initially evaluating workplace-based assessment reliability. This statistic, accounting for absolute as opposed to relative error variance, is a more robust measure of assessment reliability in this context. Again, the large number of trainers, trainees and submitted assessments is a strength of this study.

The validation of the Objective Surgical Skill examination followed a contemporary validity evidence framework (Messick's) and benefitted from a prospective approach. The naturalistic observation of an implemented SBA as it occurs in practice allows us to investigate the validity of SBA in a context-specific fashion. Again, generalizability theory was used to compare the reliability of workplace-based and simulation-based assessments of technical skill. Since the OSS examination is used to rank candidates (and contributes toward decisions on progression to Higher Specialist Training), the 'G' statistic is a more appropriate measure of reliability than the index of dependability Φ.

In Chapter 5, we discuss the development of a novel simulation-based assessment of operative competence for use in Higher Specialist Training in General Surgery. A strength of this chapter is the formalized approach to content validity generation by producing a prioritised procedural list for use in station-generation. Though a Copenhagen Academy for Medical Education and Simulation (CAMES) Needs Assessment Formula has been used to generate content for training to a simulation-based curriculum[645-647], this is the first time (to our knowledge) that a modified

version of this formula has been used to generate content validity evidence for trainee assessment.

Our novel assessment was further developed and evaluated using a contemporary approach as outlined by Cook and Hatala (2016)[173]. Our mixed-method approach to validity evidence generation, incorporating qualitative methodology, sheds new light on the use of simulation-based assessment in senior training. Evidence regarding the confounding effect of utilized models, and the perceived inability of simulation to capture contextual competence, are two important concepts which threaten the validity of simulation as an assessment modality and would not have been recorded by quantitative methods alone.

*6.5.2 Limitations*

The heterogeneity of methods, tools and assessment contexts reported in the medical education literature prevented meta-analysis in the reviews reported as part of this thesis. In Chapter 2, individual representatives from stakeholder groups were recruited to describe their perceptions and experiences of simulation-based assessment. This generated a diverse range of views, though critics of this approach may highlight that the heterogeneity of such a group limits the conclusions that can be drawn. Furthermore, the inclusion of a simulation-industry representative could be viewed as introducing bias, though in our opinion the addition of industry knowledge was important to provide a perspective on the feasibility of - and technological considerations related to - simulation as an assessment modality. The response rate for our trainee survey was poor. In hindsight, this may be due to the large number of items contained within this survey. A shorter survey tool may have yielded a higher response rate.

While in chapter three we have explored the validity of technical aptitude assessments in surgical trainee selection, we recognize that technical aptitude is just one of several domains that are valued at the time of recruitment. It is difficult, if not impossible, for any selection framework to account for such a diverse range of

sought after technical and non-technical skills and attributes. The concept of 'tests predicting tests', where candidates who are proficient at undertaking assessments continue to excel across assessments of similar constructs across training, is a threat to assessment validity at the time of recruitment. The extent to which this concept impacts upon validity again relates to the extent in which aptitude assessments are assessments of innate ability and not acquired skill. Furthermore, the relationship between measured factors at the time of selection and progression to Higher Specialist Training depends further on the validity of the selection process informing decisions on progression from Core Surgical training to Higher Specialist Training; this is not explored as part of this thesis.

In Chapter four we use a 'Task, Gap' Action' framework to evaluate the feedback provided to trainees as part of workplace-based assessments. As the 'task' at hand is readily apparent from the procedure being assessed, this domain was not analysed for submitted assessments, and thus the tool is *modified*. The validity of this modified tool has not been assessed. We further extensively use generalizability theory to explore the reliability of assessments in this chapter. In theory, a wide range of potential sources of measurement error can be analysed using this theory and related methods. This is somewhat limited by what can be, and has been, assessed by the workplace-based and simulation-based assessment tools used. The comparison of reliability metrics across assessment modalities (i.e. comparing the reliability of simulation- and workplace-based assessment) generates further methodological challenges. An important consideration would be whether workplace-based and simulation-based assessments in Core Surgical Training assess the same construct. As SBA at this level assesses a well-defined set of procedural skills and tasks, the construct represented by workplace-based assessments is broader, incorporating *context-specific* competencies that can only be assessed in the workplace. While certainly a limitation when comparing reliability metrics, this concept further supports our hypothesis that simulation and workplace-based assessments play a complimentary role in surgical training.

In Chapter five, an important consideration is whether our approach to content generation (using a Copenhagen Academy for Medical Education and Simulation [CAMES] needs assessment framework) represents a true 'Delphi' process. While initial content was generated from the existing general surgery training curriculum of Ireland and the United Kingdom (and not initial brainstorming) such an approach was deemed to add value to this study given its aim to establish content for an assessment designed to align with this established curriculum. Feasibility and cost considerations limited the number of iterations of our novel assessment that we could run, and this in turn impacted upon the sample size. Furthermore, comparing scores awarded to higher specialist trainees of closer grade, or comparing senior trainees to scores achieved by consultant colleagues, may have added to the methodological rigor of our approach. Recruitment and cost challenges prevented further exploration in this direction. A larger sample of trainers and trainees recruited to our qualitative case study may have allowed for a 'grounded theory' approach to theory generation regarding the acceptable use of SBA to inform certification decisions. Such an approach may be considered in future.

The overarching framework used to develop the research questions for this thesis was Van der Vleuten's utility formula[22], which posits that the usefulness of an assessment is a function of its validity, reliability, cost effectiveness, acceptability and educational impact. It is clear that much work remains to be done to evaluate each of these domains across the assessment time-points reported in this thesis. At selection level, the reliability of technical aptitude assessments can and should be formally investigated. Furthermore, the *social validity,* related to assessment acceptability, of these aptitude assessments could be further investigated. Future prospective work will be required to ensure that RCSI's utilized selection framework meets high standards of diversity, equity and inclusion. Across our reported assessments, and indeed across the assessment literature, there is a lack of evidence regarding cost-effectiveness. How the cost-effectiveness of such assessments should be reported and measured is an area for future study development.

## 6.6 Conclusion

In conclusion, this thesis presents utility evidence for the use of high-stakes technical performance assessments across the continuum of surgical training, from trainee selection to certification. We have identified strengths and limitations of currently utilized assessment modalities and have suggested areas where simulation or aptitude assessments can improve the validity of programmatic assessment. The validity evidence of technical aptitude assessments were explored at the time of trainee selection, with evidence to date suggesting that such assessments can predict technical performance in early surgical training. We have further explored threats to the validity of currently utilized workplace-based assessments in Core Surgical Training and have prospectively collated validity evidence for an established simulation-based assessment of technical skill at this level. An evidence- and theory-based approach to assessment development was used to design a novel simulation-based assessment intended to examine the operative competence of senior surgical trainees, with acceptable quantitative validity metrics generated. Qualitative evidence generated from our work has suggested a framework by which we should design technical performance assessments across surgical training. Across the continuum, an increasing number of domains contribute to 'competence'. In order to assess these domains meaningfully, assessments of increasing professional authenticity are required over the course of surgical training. Simulation-based assessments may struggle to reach professional authenticity levels approaching that of workplace-based assessment, which limits the ability of such assessments to assess *contextual* competence in surgical trainees. We therefore envisage a complimentary role of simulation-based assessment across surgical training. It can be acceptably used early in training, prior to in-person operating, to ensure canonical competence in core procedural skills and tasks. In more senior training, it can again be used to assess canonical competence in procedures encountered too infrequently in practice to permit safe and reliable workplace-based assessment.

## 7. Bibliography


1. Hurreiz H. The evolution of surgical training in the UK. *Adv Med Educ Pract* 2019;10:163-68. doi: 10.2147/AMEP.S189298
2. Brown E, Choi J, Sairi T. Resident Involvement in Plastic Surgery: Divergence of Patient Expectations and Experiences with Surgeon's Attitudes and Actions. *Journal of surgical education* 2020;77(2):291-99. doi: https://doi.org/10.1016/j.jsurg.2019.10.008
3. Aggarwal R, Darzi A. Training in the operating theatre: is it safe? *Thorax* 2006;61(4):278-79. doi: 10.1136/thx.2005.051789
4. Elfenbein DM. Confidence Crisis Among General Surgery Residents: A Systematic Review and Qualitative Discourse Analysis. *JAMA surgery* 2016;151(12):1166-75. doi: 10.1001/jamasurg.2016.2792
5. Poudel S, Hirano S, Kurashima Y, et al. Are graduating residents sufficiently competent? Results of a national gap analysis survey of program directors and graduating residents in Japan. *Surgery Today* 2020;50(9):995-1001. doi: 10.1007/s00595-020-01981-0
6. Vu JV, George BC, Clark M, et al. Readiness of Graduating General Surgery Residents to Perform Colorectal Procedures. *Journal of Surgical Education* 2021 doi: https://doi.org/10.1016/j.jsurg.2020.12.015
7. Harris KA, Nousiainen MT, Reznick R. Competency-based resident education-The Canadian perspective. *Surgery* 2020;167(4):681-84. doi: 10.1016/j.surg.2019.06.033 [published Online First: 2019/08/23]
8. The general surgery milestone project. *J Grad Med Educ* 2014;6(1 Suppl 1):320-28. doi: 10.4300/JGME-06-01s1-40.1
9. James HK, Gregory RJH. The dawn of a new competency-based training era. *Bone & Joint Open* 2021;2(3):181-90. doi: 10.1302/2633-1462.23.BJO-2020-0155.R1
10. Schijven MP, Bemelman WA. Problems and pitfalls in modern competency-based laparoscopic Training. *Surg Endosc* 2011;25(7):2159-63. doi: 10.1007/s00464-010-1515-y
11. Cook TL, J. . General Surgery Curriculum - The Intercollegiate Surgical Curriculum Programme. 2021 2020. https://www.iscp.ac.uk/media/1103/general-surgery-curriculum-aug-2021-approved-oct-20v3.pdf (accessed 03/02/2021).
12. Agha RA, Fowler AJ. The role and validity of surgical simulation. *International surgery* 2015;100(2):350-57. doi: 10.9738/INTSURG-D-14-00004.1
13. Milburn JA, Khera G, Hornby ST, et al. Introduction, availability and role of simulation in surgical education and training: Review of current evidence and recommendations from the Association of Surgeons in Training. *International Journal of Surgery* 2012;10(8):393-98. doi: https://doi.org/10.1016/j.ijsu.2012.05.005
14. de Montbrun SL, Roberts PL, Lowry AC, et al. A novel approach to assessing technical competence of colorectal surgery residents: the development and evaluation of the Colorectal Objective Structured Assessment of Technical Skill (COSATS). *Ann Surg* 2013;258(6):1001-6. doi: 10.1097/SLA.0b013e31829b32b8 [published Online First: 2013/07/03]
15. Halwani Y, Sachdeva AK, Satterthwaite L, et al. Development and evaluation of the General Surgery Objective Structured Assessment of Technical Skill (GOSATS). *Br J Surg* 2019;106(12):1617-22. doi: 10.1002/bjs.11359 [published Online First: 2019/10/08]

16. Buckley CE, Kavanagh DO, Traynor O, et al. Is the skillset obtained in surgical simulation transferable to the operating theatre? *Am J Surg* 2014;207(1):146-57. doi: 10.1016/j.amjsurg.2013.06.017 [published Online First: 2013/11/19]
17. Meling TR, Meling TR. The impact of surgical simulation on patient outcomes: a systematic review and meta-analysis. *Neurosurgical Review* 2021;44(2):843-54. doi: 10.1007/s10143-020-01314-2
18. de Montbrun S, Roberts PL, Satterthwaite L, et al. Implementing and Evaluating a National Certification Technical Skills Examination: The Colorectal Objective Structured Assessment of Technical Skill. *Ann Surg* 2016;264(1):1-6. doi: 10.1097/sla.0000000000001620 [published Online First: 2016/01/15]
19. Dupre J, Naik VN. The role of simulation in high-stakes assessment. *BJA Educ* 2021;21(4):148-53. doi: 10.1016/j.bjae.2020.12.002 [published Online First: 2021/03/30]
20. Cate Ot. Entrustment as Assessment: Recognizing the Ability, the Right, and the Duty to Act. *Journal of Graduate Medical Education* 2016;8(2):261-62. doi: 10.4300/jgme-d-16-00097.1
21. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery* 2021;88:105906. doi: https://doi.org/10.1016/j.ijsu.2021.105906
22. van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ* 2005;39(3):309-17. doi: 10.1111/j.1365-2929.2005.02094.x [published Online First: 2005/03/01]
23. National Academies of Sciences E, Medicine, Health, et al. Training the Future Child Health Care Workforce to Improve the Behavioral Health of Children, Youth, and Families: Proceedings of a Workshop. Washington (DC): National Academies Press (US)
Copyright 2017 by the National Academy of Sciences. All rights reserved. 2017.
24. Reed DA, Cook DA, Beckman TJ, et al. Association between funding and quality of published medical education research. *Jama* 2007;298(9):1002-9. doi: 10.1001/jama.298.9.1002 [published Online First: 2007/09/06]
25. Winkler-Schwartz A, Bajunaid K, Mullah MAS, et al. Bimanual Psychomotor Performance in Neurosurgical Resident Applicants Assessed Using NeuroTouch, a Virtual Reality Simulator. *J Surg Educ* 2016;73(6):942-53. doi: 10.1016/j.jsurg.2016.04.013 [published Online First: 2016/07/11]
26. Buckley CE, Kavanagh DO, Nugent E, et al. The impact of aptitude on the learning curve for laparoscopic suturing. *The American Journal of Surgery* 2014;207(2):263-70.
27. Gangemi A, Shi K, Durgam S, et al. A Call to Evaluate Manual Dexterity of Prospective Surgical Trainees. *J Surg Res* 2022;279:518-25. doi: 10.1016/j.jss.2022.06.025 [published Online First: 2022/07/23]
28. Groenier M, Groenier KH, Miedema HAT, et al. Perceptual Speed and Psychomotor Ability Predict Laparoscopic Skill Acquisition on a Simulator. *Journal of Surgical Education* 2015;72(6):1224-32. doi: https://doi.org/10.1016/j.jsurg.2015.07.006
29. Toale C, Morris M, Ryan DM, et al. Baseline Assessments of Psychomotor Aptitude are Associated with the Future Operative Performance of Surgical Trainees A Prospective Observational Cohort Study. *Ann Surg* 2022 doi: 10.1097/sla.0000000000005565 [published Online First: 2022/07/16]

30. Goldberg AE, Neifeld JP, Wolfe LG, et al. Correlation of manual dexterity with USMLE scores and medical student class rank. *J Surg Res* 2008;147(2):212-5. doi: 10.1016/j.jss.2008.02.050 [published Online First: 2008/05/24]
31. Moglia A, Morelli L, Ferrari V, et al. Distribution of innate psychomotor skills recognized as important for surgical specialization in unconditioned medical undergraduates. *Surg Endosc* 2018;32(10):4087-95. doi: 10.1007/s00464-018-6146-8 [published Online First: 2018/03/16]
32. Moglia A, Ferrari V, Morelli L, et al. Distribution of innate ability for surgery amongst medical students assessed by an advanced virtual reality surgical simulator. *Surg Endosc* 2014;28(6):1830-7. doi: 10.1007/s00464-013-3393-6 [published Online First: 2014/01/21]
33. Louridas M, Szasz P, Fecso AB, et al. Practice does not always make perfect: need for selection curricula in modern surgical training. *Surg Endosc* 2017;31(9):3718-27. doi: 10.1007/s00464-017-5572-3 [published Online First: 2017/04/30]
34. Cope DH, Fenton-Lee D. Assessment of laparoscopic psychomotor skills in interns using the MIST Virtual Reality Simulator: a prerequisite for those considering surgical training? *ANZ J Surg* 2008;78(4):291-6. doi: 10.1111/j.1445-2197.2007.04440.x [published Online First: 2008/03/28]
35. Panait L, Larios JM, Brenes RA, et al. Surgical skills assessment of applicants to general surgery residency. *J Surg Res* 2011;170(2):189-94. doi: 10.1016/j.jss.2011.04.006 [published Online First: 2011/05/27]
36. Lund S, D'Angelo JD, Baloul M, et al. Simulation as Soothsayer: Simulated Surgical Skills MMIs During Residency Interviews are Associated With First Year Residency Performance. *J Surg Educ* 2022;79(6):e235-e41. doi: 10.1016/j.jsurg.2022.06.002 [published Online First: 2022/06/21]
37. Louridas M, Szasz P, de Montbrun S, et al. International assessment practices along the continuum of surgical training. *Am J Surg* 2016;212(2):354-60. doi: 10.1016/j.amjsurg.2015.12.017 [published Online First: 2016/03/29]
38. AlJamal Y, Prabhakar N, Rivera M, et al. A simulation-based selection process for trying to identify medical students who will become outstanding general surgery residents. *Am J Surg* 2020;220(1):76-82. doi: 10.1016/j.amjsurg.2019.10.050 [published Online First: 2019/11/19]
39. Masood MM, Stephenson ED, Farquhar DR, et al. Surgical simulation and applicant perception in otolaryngology residency interviews. *Laryngoscope* 2018;128(11):2503-07. doi: 10.1002/lary.27211 [published Online First: 2018/04/27]
40. Toale C, Morris M, Kavanagh DO. Perspectives on simulation-based assessment of operative skill in surgical training. *Med Teach* 2022:1-8. doi: 10.1080/0142159x.2022.2134001 [published Online First: 2022/10/27]
41. Gardner AK, Steffes CP, Nepomnayshy D, et al. Selection bias: Examining the feasibility, utility, and participant receptivity to incorporating simulation into the general surgery residency selection process. *Am J Surg* 2017;213(6):1171-77. doi: 10.1016/j.amjsurg.2016.09.029 [published Online First: 2016/12/29]
42. Gillis ME, Scott SA, Richardson CG, et al. Developing and Assessing the Feasibility of Implementing a Surgical Objective Structured Clinical Skills Examination (S-OSCE). *J Surg Educ* 2020;77(4):939-46. doi: 10.1016/j.jsurg.2020.02.018 [published Online First: 2020/03/18]

43. Gallagher AG, Neary P, Gillen P, et al. Novel method for assessment and selection of trainees for higher surgical training in general surgery. *ANZ J Surg* 2008;78(4):282-90. doi: 10.1111/j.1445-2197.2008.04439.x [published Online First: 2008/03/28]
44. Carroll SM, Kennedy AM, Traynor O, et al. Objective assessment of surgical performance and its impact on a national selection programme of candidates for higher surgical training in plastic surgery. *J Plast Reconstr Aesthet Surg* 2009;62(12):1543-9. doi: 10.1016/j.bjps.2008.06.054 [published Online First: 2008/10/22]
45. Gallagher AG, O'Sullivan GC, Neary PC, et al. An objective evaluation of a multi-component, competitive, selection process for admitting surgeons into higher surgical training in a national setting. *World J Surg* 2014;38(2):296-304. doi: 10.1007/s00268-013-2302-4 [published Online First: 2013/10/23]
46. Tansley P, Kakar S, Withey S, et al. Visuospatial and technical ability in the selection and assessment of higher surgical trainees in the London deanery. *Ann R Coll Surg Engl* 2007;89(6):591-5. doi: 10.1308/003588407x187702 [published Online First: 2008/01/19]
47. Hazboun R, Rodriguez S, Thirumoorthi A, et al. Laparoscopic skills assessment: an additional modality for pediatric surgery fellowship selection. *J Pediatr Surg* 2017;52(12):1904-08. doi: 10.1016/j.jpedsurg.2017.08.051 [published Online First: 2017/09/28]
48. Erturan G, Alvand A, Judge A, et al. Prior Generic Arthroscopic Volume Correlates with Hip Arthroscopic Proficiency: A Simulator Study. *J Bone Joint Surg Am* 2018;100(1):e3. doi: 10.2106/jbjs.17.00352 [published Online First: 2018/01/04]
49. Salgado J, Grantcharov TP, Papasavas PK, et al. Technical skills assessment as part of the selection process for a fellowship in minimally invasive surgery. *Surg Endosc* 2009;23(3):641-4. doi: 10.1007/s00464-008-0033-7 [published Online First: 2008/09/25]
50. Gearhart SL, Wang MH, Gilson MM, et al. Teaching and assessing technical proficiency in surgical subspecialty fellowships. *J Surg Educ* 2012;69(4):521-8. doi: 10.1016/j.jsurg.2012.04.004 [published Online First: 2012/06/09]
51. Burden C, Fox R, Lenguerrand E, et al. Curriculum development for basic gynaecological laparoscopy with comparison of expert trainee opinions; prospective cross-sectional observational study. *Eur J Obstet Gynecol Reprod Biol* 2014;180:1-7. doi: 10.1016/j.ejogrb.2014.05.036 [published Online First: 2014/06/29]
52. Doneza JA, Palvia V, Lerner VT, et al. Correlation of surgical case volume and fellowship training with performance on simulated procedural tasks. *American Journal of Obstetrics and Gynecology* 2021;225(5):548.e1-48.e10. doi: https://doi.org/10.1016/j.ajog.2021.06.062
53. Goderstad JM, Fosse E, Sandvik L, et al. Development and validation of a curriculum for laparoscopic supracervical hysterectomy. *Facts Views Vis Obgyn* 2020;12(2):83-90. [published Online First: 2020/08/25]
54. de Vries AH, Muijtjens AMM, van Genugten HGJ, et al. Development and validation of the TOCO-TURBT tool: a summative assessment tool that measures surgical competency in transurethral resection of bladder tumour. *Surg Endosc* 2018;32(12):4923-31. doi: 10.1007/s00464-018-6251-8 [published Online First: 2018/06/07]
55. Bube SH, Hansen RB, Dahl C, et al. Development and validation of a simulator-based test in transurethral resection of bladder tumours (TURBEST). *Scand J Urol* 2019;53(5):319-24. doi: 10.1080/21681805.2019.1663921 [published Online First: 2019/09/21]

56. Aljamal YN, Farley DR. Making EPAs a 59 minute objective measure for surgical trainees - A pilot study. *Am J Surg* 2020;219(1):185-90. doi: 10.1016/j.amjsurg.2019.07.034 [published Online First: 2019/08/05]
57. Haidari TA, Bjerrum F, Christensen TD, et al. Assessing VATS competence based on simulated lobectomies of all five lung lobes. *Surg Endosc* 2022;36(11):8067-75. doi: 10.1007/s00464-022-09235-5 [published Online First: 2022/04/26]
58. Petersen RH, Gjeraa K, Jensen K, et al. Assessment of competence in video-assisted thoracoscopic surgery lobectomy: A Danish nationwide study. *J Thorac Cardiovasc Surg* 2018;156(4):1717-22. doi: 10.1016/j.jtcvs.2018.04.046 [published Online First: 2018/05/19]
59. Reliable And Valid Assessment Of Thoracoscopy Performance. D41 DEVELOPING AND IMPROVING MEASURES TO ASSESS AND CHANGE BELIEFS, SKILLS, BEHAVIORS AND OUTCOMES:A5768-A68.
60. Engberg M, Lönn L, Konge L, et al. Reliable and valid assessment of procedural skills in resuscitative endovascular balloon occlusion of the aorta. *J Trauma Acute Care Surg* 2021;91(4):663-71. doi: 10.1097/ta.0000000000003338 [published Online First: 2021/07/06]
61. Thomsen ASS, Kiilgaard JF, Kjaerbo H, et al. Simulation-based certification for cataract surgery. *Acta Ophthalmol* 2015;93(5):416-21. doi: 10.1111/aos.12691 [published Online First: 2015/02/28]
62. Jacobsen ME, Andersen MJ, Hansen CO, et al. Testing basic competency in knee arthroscopy using a virtual reality simulator: exploring validity and reliability. *J Bone Joint Surg Am* 2015;97(9):775-81. doi: 10.2106/jbjs.N.00747 [published Online First: 2015/05/08]
63. Dwyer T, Wadey V, Archibald D, et al. Cognitive and Psychomotor Entrustable Professional Activities: Can Simulators Help Assess Competency in Trainees? *Clin Orthop Relat Res* 2016;474(4):926-34. doi: 10.1007/s11999-015-4553-x [published Online First: 2015/09/24]
64. Nayahangan LJ, Lawaetz J, Strøm M, et al. Ensuring Competency in Open Aortic Aneurysm Repair – Development and Validation of a New Assessment Tool. *European Journal of Vascular and Endovascular Surgery* 2020;59(5):767-74. doi: https://doi.org/10.1016/j.ejvs.2020.01.021
65. Thinggaard E, Bjerrum F, Strandbygaard J, et al. Ensuring Competency of Novice Laparoscopic Surgeons—Exploring Standard Setting Methods and their Consequences. *Journal of Surgical Education* 2016;73(6):986-91. doi: https://doi.org/10.1016/j.jsurg.2016.05.008
66. Tedesco MM, Pak JJ, Harris EJ, Jr., et al. Simulation-based endovascular skills assessment: the future of credentialing? *J Vasc Surg* 2008;47(5):1008-1; discussion 14. doi: 10.1016/j.jvs.2008.01.007 [published Online First: 2008/03/29]
67. Culligan P, Gurshumov E, Lewis C, et al. Predictive validity of a training protocol using a robotic surgery simulator. *Female Pelvic Med Reconstr Surg* 2014;20(1):48-51. doi: 10.1097/spv.0000000000000045 [published Online First: 2013/12/26]
68. Deutsch ES, Wiet GJ, Seidman M, et al. Simulation Activity in Otolaryngology Residencies. *Otolaryngol Head Neck Surg* 2015;153(2):193-201. doi: 10.1177/0194599815584598 [published Online First: 2015/05/29]
69. Janssens S, Beckmann M, Bonney D. Introducing a laparoscopic simulation training and credentialing program in gynaecology: an observational study. *Aust N Z J Obstet*

*Gynaecol* 2015;55(4):374-8. doi: 10.1111/ajo.12355 [published Online First: 2015/07/16]
70. Ahmed K, Khan R, Mottrie A, et al. Development of a standardised training curriculum for robotic surgery: a consensus statement from an international multidisciplinary group of experts. *BJU Int* 2015;116(1):93-101. doi: 10.1111/bju.12974 [published Online First: 2014/11/02]
71. Awad M, Awad F, Carter F, et al. Consensus views on the optimum training curriculum for advanced minimally invasive surgery: A delphi study. *Int J Surg* 2018;53:137-42. doi: 10.1016/j.ijsu.2018.03.039 [published Online First: 2018/03/28]
72. Seaman SJ, Jorgensen EM, Tramontano AC, et al. Use of Fundamentals of Laparoscopic Surgery Testing to Assess Gynecologic Surgeons: A Retrospective Cohort Study of 10-Years Experience. *J Minim Invasive Gynecol* 2021;28(4):794-800. doi: 10.1016/j.jmig.2020.07.005 [published Online First: 2020/07/19]
73. Bilgic E, Kaneva P, Okrainec A, et al. Trends in the Fundamentals of Laparoscopic Surgery® (FLS) certification exam over the past 9 years. *Surg Endosc* 2018;32(4):2101-05. doi: 10.1007/s00464-017-5907-0 [published Online First: 2017/10/27]
74. Rooney DM, Brissman IC, Gauger PG. Ongoing evaluation of video-based assessment of proctors' scoring of the fundamentals of laparoscopic surgery manual skills examination. *J Surg Educ* 2015;72(3):471-6. doi: 10.1016/j.jsurg.2014.10.017 [published Online First: 2015/04/13]
75. Ritter EM, Brissman IC. Systematic development of a proctor certification examination for the Fundamentals of Laparoscopic Surgery testing program. *Am J Surg* 2016;211(2):458-63. doi: 10.1016/j.amjsurg.2015.10.001 [published Online First: 2015/12/15]
76. Xeroulis G, Dubrowski A, Leslie K. Simulation in laparoscopic surgery: a concurrent validity study for FLS. *Surg Endosc* 2009;23(1):161-5. doi: 10.1007/s00464-008-0120-9 [published Online First: 2008/09/25]
77. Fong ML, Leeder P, Dexter S, et al. Development and evaluation of an objective laparoscopic assessment tool: LapPass®. *Ann R Coll Surg Engl* 2022;104(5):367-72. doi: 10.1308/rcsann.2021.0207 [published Online First: 2021/11/26]
78. Duran C, Estrada S, O'Malley M, et al. The model for Fundamentals of Endovascular Surgery (FEVS) successfully defines the competent endovascular surgeon. *J Vasc Surg* 2015;62(6):1660-6.e3. doi: 10.1016/j.jvs.2015.09.026 [published Online First: 2015/11/26]
79. Belvroy VM, Murali B, Sheahan MG, et al. In the Fundamentals of Endovascular and Vascular Surgery model motion metrics reliably differentiate competency. *Journal of Vascular Surgery* 2020;72(6):2161-65. doi: https://doi.org/10.1016/j.jvs.2020.02.047
80. Satava R, Gallagher AG. Proficiency-based progression process training for fundamentals of robotic surgery curriculum development. *Annals of Laparoscopic and Endoscopic Surgery* 2020;5
81. Satava RM, Stefanidis D, Levy JS, et al. Proving the Effectiveness of the Fundamentals of Robotic Surgery (FRS) Skills Curriculum: A Single-blinded, Multispecialty, Multi-institutional Randomized Control Trial. *Ann Surg* 2020;272(2):384-92. doi: 10.1097/sla.0000000000003220 [published Online First: 2020/07/18]
82. Martin JR, Stefanidis D, Dorin RP, et al. Demonstrating the effectiveness of the fundamentals of robotic surgery (FRS) curriculum on the RobotiX Mentor Virtual

Reality Simulation Platform. *J Robot Surg* 2021;15(2):187-93. doi: 10.1007/s11701-020-01085-4 [published Online First: 2020/05/16]
83. Sousa J, Mansilha A. European Panorama on Vascular Surgery: Results from Five Years of FEBVS Examinations. *European Journal of Vascular and Endovascular Surgery* 2021;61(5):S41-S42. doi: 10.1016/j.ejvs.2021.01.034
84. Pandey VA, Wolfe JH, Liapis CD, et al. The examination assessment of technical competence in vascular surgery. *Br J Surg* 2006;93(9):1132-8. doi: 10.1002/bjs.5302 [published Online First: 2006/07/25]
85. Halwani Y, Sachdeva AK, Satterthwaite L, et al. Development and evaluation of the General Surgery Objective Structured Assessment of Technical Skill (GOSATS). *British Journal of Surgery* 2019;106(12):1617-22. doi: 10.1002/bjs.11359
86. de Montbrun S, Louridas M, Szasz P, et al. Developing the Blueprint for a General Surgery Technical Skills Certification Examination: A Validation Study. *J Surg Educ* 2018;75(2):344-50. doi: 10.1016/j.jsurg.2017.08.007 [published Online First: 2017/09/03]
87. Stephens EH, Dearani JA. Commentary: Surgical skill assessment: Time to examine? *J Thorac Cardiovasc Surg* 2020;160(1):242-43. doi: 10.1016/j.jtcvs.2020.01.004 [published Online First: 2020/02/09]
88. Morse CR, Mathisen DJ. Educational Challenges of the Operating Room. *Thorac Surg Clin* 2019;29(3):269-77. doi: 10.1016/j.thorsurg.2019.03.005 [published Online First: 2019/06/27]
89. Hermsen JL, Mohamadipanah H, Yang S, et al. Multimodal Cardiopulmonary Bypass Skills Assessment Within a High-Fidelity Simulation Environment. *Ann Thorac Surg* 2021;112(2):652-60. doi: 10.1016/j.athoracsur.2020.07.022 [published Online First: 2020/09/25]
90. Giannotti D, Patrizi G, Casella G, et al. Can virtual reality simulators be a certification tool for bariatric surgeons? *Surg Endosc* 2014;28(1):242-8. doi: 10.1007/s00464-013-3179-x [published Online First: 2013/09/03]
91. Stefanidis D, Arora S, Parrack DM, et al. Research priorities in surgical simulation for the 21st century. *Am J Surg* 2012;203(1):49-53. doi: 10.1016/j.amjsurg.2011.05.008 [published Online First: 2011/12/17]
92. Duran C, Bismuth J, Mitchell E. A nationwide survey of vascular surgery trainees reveals trends in operative experience, confidence, and attitudes about simulation. *J Vasc Surg* 2013;58(2):524-8. doi: 10.1016/j.jvs.2012.12.072 [published Online First: 2013/04/02]
93. Awad M, Awad F, Carter F, et al. Consensus views on the optimum training curriculum for advanced minimally invasive surgery: A delphi study. *International Journal of Surgery* 2018;53:137-42. doi: https://doi.org/10.1016/j.ijsu.2018.03.039
94. Alvand A, Auplish S, Khan T, et al. Identifying orthopaedic surgeons of the future: the inability of some medical students to achieve competence in basic arthroscopic tasks despite training: a randomised study. *J Bone Joint Surg Br* 2011;93(12):1586-91. doi: 10.1302/0301-620x.93b12.27946 [published Online First: 2011/12/14]
95. Karani R, Tapiero S, Jefferson FA, et al. Crowd-Sourced Assessment of Surgical Skills of Urology Resident Applicants: Four-Year Experience. *J Surg Educ* 2021;78(6):2030-37. doi: 10.1016/j.jsurg.2021.05.005 [published Online First: 2021/06/21]
96. Morris MC, Frodl T, D'Souza A, et al. Assessment of competence in surgical skills using functional magnetic resonance imaging: a feasibility study. *J Surg Educ*

2015;72(2):198-204. doi: 10.1016/j.jsurg.2014.09.007 [published Online First: 2014/12/03]
97. Zahid A, Rajan V, Hong J, et al. Surgical Competencies Required in Newly Commencing Colorectal Surgeons: an Educational and Training Spectrum. *Med Sci Educ* 2020;30(3):1043-47. doi: 10.1007/s40670-020-01005-z [published Online First: 2020/06/11]
98. Cousans F, Patterson F, Edwards H, et al. Evaluating the complementary roles of an SJT and academic assessment for entry into clinical practice. *Advances in Health Sciences Education* 2017;22(2):401-13. doi: 10.1007/s10459-017-9755-4
99. Gardner AK, Dunkin BJ. Evaluation of Validity Evidence for Personality, Emotional Intelligence, and Situational Judgment Tests to Identify Successful Residents. *JAMA Surg* 2018;153(5):409-16. doi: 10.1001/jamasurg.2017.5013 [published Online First: 2017/12/29]
100. Mazzone E, Puliatti S, Amato M, et al. A Systematic Review and Meta-analysis on the Impact of Proficiency-based Progression Simulation Training on Performance Outcomes. *Ann Surg* 2021;274(2):281-89. doi: 10.1097/sla.0000000000004650 [published Online First: 2021/02/26]
101. Gardner AK, Scott DJ, Hebert JC, et al. Gearing up for milestones in surgery: Will simulation play a role? *Surgery* 2015;158(5):1421-7. doi: 10.1016/j.surg.2015.03.039 [published Online First: 2015/05/28]
102. Yudkowsky R, Park YS, Downing SM. Assessment in health professions education: Routledge 2019.
103. Messick S. Validity. Educational measurement, 3rd ed: American Council on Education 1989:13-103.
104. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther* 2014;36(5):648-62. doi: 10.1016/j.clinthera.2014.04.006 [published Online First: 2014/05/09]
105. Brennan RL. Generalizability theory. New York, NY, US: Springer-Verlag Publishing 2001.
106. Craig P, Hall Dykgraaf S, Phillips C. Using the Freeth/Kirkpatrick model to evaluate interprofessional learning outcomes in a rural setting. *Focus on Health Professional Education: A Multi-Professional Journal* 2016;17:84. doi: 10.11157/fohpe.v17i1.124
107. Oslock WM, Satiani B, Way DP, et al. A contemporary reassessment of the US surgical workforce through 2050 predicts continued shortages and increased productivity demands. *The American Journal of Surgery* 2021 doi: https://doi.org/10.1016/j.amjsurg.2021.07.033
108. Giles JA. Surgical training and the European Working Time Directive: The role of informal workplace learning. *International Journal of Surgery* 2010;8(3):179-80. doi: https://doi.org/10.1016/j.ijsu.2010.01.011
109. Traynor O. Surgical training in an era of reduced working hours. *Surgeon* 2011;9 Suppl 1:S1-2. doi: 10.1016/j.surge.2011.04.003 [published Online First: 2011/10/05]
110. Liang R, Dornan T, Nestel D. Why do women leave surgical training? A qualitative and feminist study. *The Lancet* 2019;393(10171):541-49. doi: https://doi.org/10.1016/S0140-6736(18)32612-6
111. Brown C, Abdelrahman T, Patel N, et al. Operative learning curve trajectory in a cohort of surgical trainees. *The British journal of surgery* 2017;104(10):1405-11. doi: 10.1002/bjs.10584 [published Online First: 2017/07/19]

112. Randleman JB, Wolfe JD, Woodward M, et al. The resident surgeon phacoemulsification learning curve. *Arch Ophthalmol* 2007;125(9):1215-9. doi: 10.1001/archopht.125.9.1215 [published Online First: 2007/09/12]
113. Laeeq K, Lin SY, Varela DA, et al. Achievement of competency in endoscopic sinus surgery of otolaryngology residents. *Laryngoscope* 2013;123(12):2932-4. doi: 10.1002/lary.23509 [published Online First: 2013/10/15]
114. Gifford ED, Nguyen VT, Kim JJ, et al. Variation in the learning curves of general surgery residents performing arteriovenous fistulas. *J Surg Educ* 2015;72(4):761-6. doi: 10.1016/j.jsurg.2015.02.001 [published Online First: 2015/04/23]
115. Garg A, Yadav SS, Tomar V, et al. Prospective Evaluation of Learning Curve of Urology Residents for Percutaneous Nephrolithotomy. *Urology Practice* 2016;3(3):230-35. doi: https://doi.org/10.1016/j.urpr.2015.06.009
116. Bjerrum F, Thomsen ASS, Nayahangan LJ, et al. Surgical simulation: Current practices and future perspectives for technical skills training. *Medical Teacher* 2018;40(7):668-75. doi: 10.1080/0142159X.2018.1472754
117. Hosny SG, Johnston MJ, Pucher PH, et al. Barriers to the implementation and uptake of simulation-based training programs in general surgery: a multinational qualitative study. *Journal of Surgical Research* 2017;220:419-26.e2. doi: https://doi.org/10.1016/j.jss.2017.07.020
118. Brenna C, Das S. Imperfect by design: the problematic ethics of surgical training. *Journal of Medical Ethics* 2021;47(5):350-53. doi: 10.1136/medethics-2019-105837
119. Ohye RG, Jaggers JJ, Sade RM. Must Surgeons in Training Programs Allow Residents to Operate on Their Patients to Satisfy Board Requirements? *The Annals of Thoracic Surgery* 2016;101(1):18-23. doi: https://doi.org/10.1016/j.athoracsur.2015.08.049
120. Moodie JJ, Masood I, Tint N, et al. Patients' attitudes towards trainee surgeons performing cataract surgery at a teaching hospital. *Eye* 2008;22(9):1183-86. doi: 10.1038/sj.eye.6702872
121. Hutter MM, Glasgow RE, Mulvihill SJ. Does the participation of a surgical trainee adversely impact patient outcomes? A study of major pancreatic resections in California. *Surgery* 2000;128(2):286-92. doi: 10.1067/msy.2000.107416 [published Online First: 2000/08/03]
122. Shaked A, Calderom I, Durst A. Safety of surgical procedures performed by residents. *Arch Surg* 1991;126(5):559-60. doi: 10.1001/archsurg.1991.01410290031005 [published Online First: 1991/05/01]
123. Kiran RP, Ahmed Ali U, Coffey JC, et al. Impact of resident participation in surgical operations on postoperative outcomes: National Surgical Quality Improvement Program. *Ann Surg* 2012;256(3):469-75. doi: 10.1097/SLA.0b013e318265812a [published Online First: 2012/07/31]
124. Sussman EJ, Kastanis JN, Feigin W, et al. Surgical outcome for resident and attending surgeons. *Am J Surg* 1982;144(2):250-3. doi: 10.1016/0002-9610(82)90519-0 [published Online First: 1982/08/01]
125. Pariani D, Fontana S, Zetti G, et al. Laparoscopic cholecystectomy performed by residents: a retrospective study on 569 patients. *Surgery research and practice* 2014;2014:912143-43. doi: 10.1155/2014/912143 [published Online First: 2014/01/02]

126. Kanakala V, Bawa S, Gallagher P, et al. Outcome of patients in laparoscopic training courses compared to standard patients. *The Surgeon* 2010;8(3):132-35. doi: https://doi.org/10.1016/j.surge.2009.10.030
127. Renwick AA, Bokey EL, Chapuis PH, et al. Effect of supervised surgical training on outcomes after resection of colorectal cancer. *Br J Surg* 2005;92(5):631-6. doi: 10.1002/bjs.4935 [published Online First: 2005/04/06]
128. Itani KM, DePalma RG, Schifftner T, et al. Surgical resident supervision in the operating room and outcomes of care in Veterans Affairs hospitals. *Am J Surg* 2005;190(5):725-31. doi: 10.1016/j.amjsurg.2005.06.042 [published Online First: 2005/10/18]
129. Kennedy TJ, Regehr G, Baker GR, et al. Progressive independence in clinical training: a tradition worth defending? *Academic Medicine* 2005;80(10):S106-S11.
130. Pugely AJ, Gao Y, Martin CT, et al. The effect of resident participation on short-term outcomes after orthopaedic surgery. *Clin Orthop Relat Res* 2014;472(7):2290-300. doi: 10.1007/s11999-014-3567-0 [published Online First: 2014/03/25]
131. Iannuzzi JC, Chandra A, Rickles AS, et al. Resident involvement is associated with worse outcomes after major lower extremity amputation. *Journal of Vascular Surgery* 2013;58(3):827-31.e1. doi: https://doi.org/10.1016/j.jvs.2013.04.046
132. Kasotakis G, Lakha A, Sarkar B, et al. Trainee participation is associated with adverse outcomes in emergency general surgery: an analysis of the National Surgical Quality Improvement Program database. *Ann Surg* 2014;260(3):483-90; discussion 90-3. doi: 10.1097/sla.0000000000000889 [published Online First: 2014/08/15]
133. Scott DJ, Dunnington GL. The new ACS/APDS Skills Curriculum: moving the learning curve out of the operating room. *J Gastrointest Surg* 2008;12(2):213-21. doi: 10.1007/s11605-007-0357-y [published Online First: 2007/10/11]
134. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg* 2002;236(4):458-63; discussion 63-4. doi: 10.1097/00000658-200210000-00008 [published Online First: 2002/10/09]
135. Ahlberg G, Enochsson L, Gallagher AG, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *The American Journal of Surgery* 2007;193(6):797-804. doi: https://doi.org/10.1016/j.amjsurg.2006.06.050
136. Palter VN, Grantcharov TP. Development and validation of a comprehensive curriculum to teach an advanced minimally invasive procedure: a randomized controlled trial. *Ann Surg* 2012;256(1):25-32. doi: 10.1097/SLA.0b013e318258f5aa [published Online First: 2012/06/06]
137. Ahlberg G, Enochsson L, Gallagher AG, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *Am J Surg* 2007;193(6):797-804. doi: 10.1016/j.amjsurg.2006.06.050 [published Online First: 2007/05/22]
138. Angelo RL, Ryu RK, Pedowitz RA, et al. A Proficiency-Based Progression Training Curriculum Coupled With a Model Simulator Results in the Acquisition of a Superior Arthroscopic Bankart Skill Set. *Arthroscopy* 2015;31(10):1854-71. doi: 10.1016/j.arthro.2015.07.001 [published Online First: 2015/09/06]
139. Van Sickle KR, Ritter EM, Baghai M, et al. Prospective, randomized, double-blind trial of curriculum-based training for intracorporeal suturing and knot tying. *J Am Coll Surg*

2008;207(4):560-8. doi: 10.1016/j.jamcollsurg.2008.05.007 [published Online First: 2008/10/18]
140. Allen RW, Pruitt M, Taaffe KM. Effect of Resident Involvement on Operative Time and Operating Room Staffing Costs. *J Surg Educ* 2016;73(6):979-85. doi: 10.1016/j.jsurg.2016.05.014 [published Online First: 2016/06/29]
141. Advani V, Ahad S, Gonczy C, et al. Does resident involvement effect surgical times and complication rates during laparoscopic appendectomy for uncomplicated appendicitis? An analysis of 16,849 cases from the ACS-NSQIP. *Am J Surg* 2012;203(3):347-51; discussion 51-2. doi: 10.1016/j.amjsurg.2011.08.015 [published Online First: 2012/03/01]
142. Iannuzzi JC, Chandra A, Rickles AS, et al. Resident involvement is associated with worse outcomes after major lower extremity amputation. *J Vasc Surg* 2013;58(3):827-31.e1. doi: 10.1016/j.jvs.2013.04.046 [published Online First: 2013/06/19]
143. Mazzone E, Puliatti S, Amato M, et al. A Systematic Review and Meta-Analysis on the Impact of Proficiency-Based Progression Simulation Training on Performance Outcomes. *Ann Surg* 2020 doi: 10.1097/sla.0000000000004650 [published Online First: 2021/02/26]
144. Higham H, Baxendale B. To err is human: use of simulation to enhance training and patient safety in anaesthesia. *British Journal of Anaesthesia* 2017;119:i106-i14. doi: https://doi.org/10.1093/bja/aex302
145. Blumenthal D. Making medical errors into "medical treasures". *Jama* 1994;272(23):1867-8. [published Online First: 1994/12/21]
146. Slakey DP, Simms ER, Rennie KV, et al. Using simulation to improve root cause analysis of adverse surgical outcomes. *International Journal for Quality in Health Care* 2014;26(2):144-50. doi: 10.1093/intqhc/mzu011
147. Gao P, Wang C, Liu S, et al. Simulation of operating room crisis management - hypotension training for pre-clinical students. *BMC Medical Education* 2021;21(1):60. doi: 10.1186/s12909-020-02477-8
148. Kelley K, Park H. Simulation in Surgical Trauma Training. *Current Surgery Reports* 2020;8(7):13. doi: 10.1007/s40137-020-00261-7
149. Wayne DB, Didwania A, Feinglass J, et al. Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. *Chest* 2008;133(1):56-61. doi: 10.1378/chest.07-0131 [published Online First: 2007/06/19]
150. Clarke E. Virtual reality simulation-the future of orthopaedic training? A systematic review and narrative analysis. *Advances in simulation (London, England)* 2021;6(1):2-2. doi: 10.1186/s41077-020-00153-x
151. Allum W. Improving Surgical Training. *Surgery (Oxford, Oxfordshire)* 2020;38(10):596-600. doi: 10.1016/j.mpsur.2020.07.015 [published Online First: 2020/08/30]
152. Abraham RM, Singaram VS. Using deliberate practice framework to assess the quality of feedback in undergraduate clinical skills training. *BMC Medical Education* 2019;19(1):105. doi: 10.1186/s12909-019-1547-5
153. Hashimoto DA. Surgeons and Machines can Learn from Operative Video: Will the System let Them? *Ann Surg* 2021 doi: 10.1097/sla.0000000000004899 [published Online First: 2021/04/16]
154. Yin MS, Haddawy P, Suebnukarn S, et al. Toward Intelligent Tutorial Feedback in Surgical Simulation: Robust Outcome Scoring for Endodontic Surgery. Proceedings of the 21st

International Conference on Intelligent User Interfaces. Sonoma, California, USA: Association for Computing Machinery, 2016:402–06.
155. Patel A, Patel M, Lytle N, et al. Can we become better robot surgeons through simulator practice? *Surg Endosc* 2014;28(3):847-53. doi: 10.1007/s00464-013-3231-x [published Online First: 2013/10/15]
156. McIntosh C, Macario A, Flanagan B, et al. Simulation: What does it really cost? *Simulation in Healthcare* 2006;1:109. doi: 10.1097/01266021-200600120-00041
157. Maloney S, Haines T. Issues of cost-benefit and cost-effectiveness for simulation in health professions education. *Advances in Simulation* 2016;1(1):13. doi: 10.1186/s41077-016-0020-3
158. Frank JR, Mungroo R, Ahmad Y, et al. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Med Teach* 2010;32(8):631-7. doi: 10.3109/0142159x.2010.500898 [published Online First: 2010/07/29]
159. Brasel KJ, Klingensmith ME, Englander R, et al. Entrustable Professional Activities in General Surgery: Development and Implementation. *Journal of surgical education* 2019;76(5):1174-86. doi: 10.1016/j.jsurg.2019.04.003 [published Online First: 2019/04/29]
160. Yousuf Guraya S. Workplace-based Assessment; Applications and Educational Impact. *The Malaysian journal of medical sciences : MJMS* 2015;22(6):5-10. [published Online First: 2015/11/01]
161. A G, A P, J R, et al. What do surgeons and trainees think of WBAs and how do they use them? *The Bulletin of the Royal College of Surgeons of England* 2016;98(9):408-14. doi: 10.1308/rcsbull.2016.408
162. D. P, I. RD, L. SJ. Assessment of performance in orthopaedic training. *The Journal of Bone and Joint Surgery British volume* 2005;87-B(9):1187-91. doi: 10.1302/0301-620x.87b9.16434
163. Wagner JP, Lewis CE, Tillou A, et al. Use of Entrustable Professional Activities in the Assessment of Surgical Resident Competency. *JAMA surgery* 2018;153(4):335-43. doi: 10.1001/jamasurg.2017.4547
164. Norcini JJ, McKinley DW. Assessment methods in medical education. *Teaching and Teacher Education* 2007;23(3):239-50. doi: https://doi.org/10.1016/j.tate.2006.12.021
165. Merola G, Cavallaro G, Iorio O, et al. Learning curve in open inguinal hernia repair: a quality improvement multicentre study about Lichtenstein technique. *Hernia* 2020;24(3):651-59. doi: 10.1007/s10029-019-02064-x [published Online First: 2019/11/24]
166. Ahlberg G, Kruuna O, Leijonmarck C-E, et al. Is the learning curve for laparoscopic fundoplication determined by the teacher or the pupil? *The American Journal of Surgery* 2005;189(2):184-89. doi: https://doi.org/10.1016/j.amjsurg.2004.06.043
167. Hashimoto DA, Sirimanna P, Gomez ED, et al. Deliberate practice enhances quality of laparoscopic surgical performance in a randomized controlled trial: from arrested development to expert performance. *Surg Endosc* 2015;29(11):3154-62. doi: 10.1007/s00464-014-4042-4 [published Online First: 2014/12/25]
168. Buckley CE, Kavanagh DO, Nugent E, et al. The impact of aptitude on the learning curve for laparoscopic suturing. *Am J Surg* 2014;207(2):263-70. doi: 10.1016/j.amjsurg.2013.08.037 [published Online First: 2014/01/29]

169. Köckerling F. What Is the Influence of Simulation-Based Training Courses, the Learning Curve, Supervision, and Surgeon Volume on the Outcome in Hernia Repair?-A Systematic Review. *Front Surg* 2018;5:57. doi: 10.3389/fsurg.2018.00057 [published Online First: 2018/10/17]
170. Allum W, Hornby S, Khera G, et al. General Surgery Logbook Survey. *Bulletin of The Royal College of Surgeons of England* 2013;95 doi: 10.1308/147363513X13500508920095
171. Association AER. American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. 1999
172. The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation. 2018 doi: 10.4135/9781506326139
173. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Advances in Simulation* 2016;1(1):31. doi: 10.1186/s41077-016-0033-y
174. Cook DA, Lineberry M. Consequences validity evidence: evaluating the impact of educational assessments. *Academic Medicine* 2016;91(6):785-95.
175. Peters MDJ, Godfrey CM, Khalil H, et al. Guidance for conducting systematic scoping reviews. *JBI Evidence Implementation* 2015;13(3):141-46. doi: 10.1097/xeb.0000000000000050
176. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ (Clinical research ed)* 2009;339:b2535. doi: 10.1136/bmj.b2535
177. Encyclopedia of Research Design. 2010 doi: 10.4135/9781412961288
178. Royal KD, Guskey TR. On the Appropriateness of Norm- and Criterion-referenced Assessments in Medical Education. *Ear, Nose & Throat Journal* 2015;94(7):252-54. doi: 10.1177/014556131509400701
179. Liao SC, Hunt EA, Chen W. Comparison between inter-rater reliability and inter-rater agreement in performance assessment. *Annals Academy of Medicine Singapore* 2010;39(8):613.
180. Revicki D. Internal Consistency Reliability. In: Michalos AC, ed. Encyclopedia of Quality of Life and Well-Being Research. Dordrecht: Springer Netherlands 2014:3305-06.
181. Kirkpatrick D, Kirkpatrick J. Evaluating training programs: The four levels: Berrett-Koehler Publishers 2006.
182. Freeth D, Hammick M, Koppel I, et al. A critical review of evaluations of interprofessional education. *Occasional Paper No 2* 2002
183. Awad Z, Hayden L, Robson AK, et al. Reliability and validity of procedure-based assessments in otolaryngology training. *The Laryngoscope* 2015;125(6):1328-35. doi: 10.1002/lary.24983 [published Online First: 2014/10/28]
184. Beamish AJ, Johnston MJ, Harries RL, et al. Real-world use of workplace based assessments in surgical training: A UK nationwide cross-sectional exploration of trainee perspectives and consensus recommendations from the Association of Surgeons in Training. *International Journal of Surgery* 2020 doi: https://doi.org/10.1016/j.ijsu.2020.07.068
185. Manoharan G, Sharma N, Gallacher P. Competence in using the arthroscopy stack system: a national survey of orthopaedic trainees in the UK. *Ann R Coll Surg Engl* 2020;102(2):149-52. doi: 10.1308/rcsann.2019.0131 [published Online First: 2019/09/21]

186. Ramhamadany E. Orthopaedic Trainee Opinion on the Current Procedure-based Assessment Compared to its Predecessor: A UK Trauma Centre Perspective. *Cureus* 2020;12(4):e7799. doi: 10.7759/cureus.7799 [published Online First: 2020/05/08]
187. Davies RM, Hadfield-Law L, Turner PG. Development and Evaluation of a New Formative Assessment of Surgical Performance. *J Surg Educ* 2018;75(5):1309-16. doi: 10.1016/j.jsurg.2018.02.004 [published Online First: 2018/03/28]
188. Shalhoub J, Marshall DC, Ippolito K. Perspectives on procedure-based assessments: a thematic analysis of semistructured interviews with 10 UK surgical trainees. *BMJ Open* 2017;7(3):e013417. doi: 10.1136/bmjopen-2016-013417
189. Hunter AR, Baird EJ, Reed MR. Procedure-based assessments in trauma and orthopaedic training--The trainees' perspective. *Med Teach* 2015;37(5):444-9. doi: 10.3109/0142159x.2014.956055 [published Online First: 2014/09/05]
190. Marriott J, Purdie H, Crossley J, et al. Evaluation of procedure-based assessment for assessing trainees' skills in the operating theatre. *The British journal of surgery* 2011;98(3):450-7. doi: 10.1002/bjs.7342 [published Online First: 2011/01/22]
191. Sarker SK, Maciocco M, Zaman A, et al. Operative performance in laparoscopic cholecystectomy using the Procedural-Based Assessment tool. *Am J Surg* 2010;200(3):334-40. doi: 10.1016/j.amjsurg.2009.10.025 [published Online First: 2010/06/25]
192. Hopkins L, Robinson DBT, Brown C, et al. Trauma and Orthopedic Surgery Curriculum Concordance: An Operative Learning Curve Trajectory Perspective. *Journal of surgical education* 2019;76(6):1569-78. doi: https://doi.org/10.1016/j.jsurg.2019.05.009
193. De Siqueira JR, Gough MJ. Correlation between experience targets and competence for general surgery certification. *Br J Surg* 2016;103(7):921-7. doi: 10.1002/bjs.10145 [published Online First: 2016/04/10]
194. Abdelrahman T, Long J, Egan R, et al. Operative Experience vs. Competence: A Curriculum Concordance and Learning Curve Analysis. *J Surg Educ* 2016;73(4):694-8. doi: 10.1016/j.jsurg.2016.01.011 [published Online First: 2016/03/12]
195. Elsey EJ, Griffiths G, West J, et al. Changing Autonomy in Operative Experience Through UK General Surgery Training: A National Cohort Study. *Annals of surgery* 2019;269(3):399-406. doi: 10.1097/sla.0000000000003032 [published Online First: 2018/09/28]
196. Beard JD, Marriott J, Purdie H, et al. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. *Health Technol Assess* 2011;15(1):i-xxi, 1-162. doi: 10.3310/hta15010 [published Online First: 2011/01/14]
197. Gaunt A, Patel A, Rusius V, et al. 'Playing the game': How do surgical trainees seek feedback using workplace-based assessment? *Med Educ* 2017;51(9):953-62. doi: 10.1111/medu.13380 [published Online First: 2017/08/24]
198. Eardley I, Bussey M, Woodthorpe A, et al. Workplace-based assessment in surgical training: experiences from the Intercollegiate Surgical Curriculum Programme. *ANZ J Surg* 2013;83(6):448-53. doi: 10.1111/ans.12187 [published Online First: 2013/05/10]
199. Awad Z, Hayden L, Muthuswamy K, et al. Does direct observation of procedural skills reflect trainee's progress in otolaryngology? *Clin Otolaryngol* 2014;39(3):169-73. doi: 10.1111/coa.12251 [published Online First: 2014/05/08]
200. Beamish AJ, Johnston MJ, Harries RL, et al. Use of the eLogbook in surgical training in the United Kingdom: A nationwide survey and consensus recommendations from the

Association of Surgeons in Training. *International Journal of Surgery* 2020 doi: https://doi.org/10.1016/j.ijsu.2020.02.045
201. Aryal KR, Currow C, Downey S, et al. Work-Based Assessments in Higher General Surgical Training Program: A Mixed Methods Study Exploring Trainers' and Trainees' Views and Experiences. *Surg J (N Y)* 2020;6(1):e49-e61. doi: 10.1055/s-0040-1708062 [published Online First: 2020/03/12]
202. Gaunt A, Patel A, Royle T, et al. What do surgeons and trainees think of WBAs and how do they use them? *The Bulletin of the Royal College of Surgeons of England* 2016;98:408-14. doi: 10.1308/rcsbull.2016.408
203. Pereira EA, Dean BJ. British surgeons' experiences of a mandatory online workplace based assessment portfolio resurveyed three years on. *Journal of surgical education* 2013;70(1):59-67. doi: 10.1016/j.jsurg.2012.06.019 [published Online First: 2013/01/23]
204. Roushdi I, Tennent D. Current usage patterns of procedure-based assessments in the orthopaedic community. *Bulletin of The Royal College of Surgeons of England* 2015;97:e1-e3. doi: 10.1308/147363515X14134529300148
205. Satpathy A, Kneebone R. Workplace-Based Assessment Insurgical Training: A Qualitative Study. *Bulletin of The Royal College of Surgeons of England* 2011;93:1-5. doi: 10.1308/147363511X595082
206. Pereira EAC, Dean BJF. British surgeons' experiences of mandatory online workplace-based assessment. *J R Soc Med* 2009;102(7):287-93. doi: 10.1258/jrsm.2009.080398
207. Fu B, Amin K, Hili S, et al. Assessing ISCP users' satisfaction in ENT trainers and trainees. *Clinical otolaryngology : official journal of ENT-UK ; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery* 2011;36(6):594-6. doi: 10.1111/j.1749-4486.2011.02409.x [published Online First: 2012/01/04]
208. Phillips AW, Madhavan A, Bookless LR, et al. Surgical Trainers' Experience and Perspectives on Workplace-Based Assessments. *Journal of surgical education* 2015;72(5):979-84. doi: https://doi.org/10.1016/j.jsurg.2015.03.015
209. Basu I, Parvizi S, Chin K. The perception of online work-based assessments. *The clinical teacher* 2013;10(2):73-7. doi: 10.1111/j.1743-498X.2012.00609.x [published Online First: 2013/03/14]
210. L. SJ, R. RM, P. C, et al. Influencing the national training agenda. *The Journal of Bone and Joint Surgery British volume* 2005;87-B(9):1182-86. doi: 10.1302/0301-620x.87b9.16433
211. Bohnen JD, George BC, Williams RG, et al. The Feasibility of Real-Time Intraoperative Performance Assessment With SIMPL (System for Improving and Measuring Procedural Learning): Early Experience From a Multi-institutional Trial. *J Surg Educ* 2016;73(6):e118-e30. doi: 10.1016/j.jsurg.2016.08.010 [published Online First: 2016/11/26]
212. Maguire T, Mayne CJ, Terry T, et al. Analysis of the surgical learning curve using the cumulative sum (CUSUM) method. *Neurourol Urodyn* 2013;32(7):964-7. doi: 10.1002/nau.22375 [published Online First: 2013/01/30]
213. Thomas C, Elsey E, Boyce T, et al. Data from two deaneries regarding the ability of higher surgical training programmes to deliver curriculum targets. *The Bulletin of the Royal College of Surgeons of England* 2015;97:e50-e53. doi: 10.1308/rcsbull.2015.e50

214. Gervais J. The operational definition of competency-based education. *The Journal of Competency-Based Education* 2016;1(2):98-106. doi: https://doi.org/10.1002/cbe2.1011
215. Hopper AN, Jamison MH, Lewis WG. Learning curves in surgical practice. *Postgrad Med J* 2007;83(986):777-79. doi: 10.1136/pgmj.2007.057190
216. Morche J, Mathes T, Pieper D. Relationship between surgeon volume and outcomes: a systematic review of systematic reviews. *Syst Rev* 2016;5(1):204. doi: 10.1186/s13643-016-0376-4 [published Online First: 2016/12/03]
217. Friedell ML, VanderMeer TJ, Cheatham ML, et al. Perceptions of graduating general surgery chief residents: are they confident in their training? *J Am Coll Surg* 2014;218(4):695-703. doi: 10.1016/j.jamcollsurg.2013.12.022 [published Online First: 2014/02/18]
218. Fonseca AL, Reddy V, Longo WE, et al. Graduating general surgery resident operative confidence: perspective from a national survey. *J Surg Res* 2014;190(2):419-28. doi: 10.1016/j.jss.2014.05.014 [published Online First: 2014/06/09]
219. McKee RF. The Intercollegiate Surgical Curriculum Programme (ISCP). *Surgery* 2008;26(10):411-16. doi: 10.1016/j.mpsur.2008.08.007
220. Wood S, James OP, Hopkins L, et al. Variations in competencies needed to complete surgical training. *BJS Open* 2019;3(6):852-56. doi: 10.1002/bjs5.50200
221. Education ACfGM. Defined Category Minimum Numbers for General Surgery Residents and Credit Role acgme.org: ACGME; 2019 [Available from: https://www.acgme.org/Portals/0/DefinedCategoryMinimumNumbersforGeneralSurgeryResidentsandCreditRole.pdf accessed 08/09/21 2021.
222. Liebman DL, McKay KM, Haviland MJ, et al. Quantifying the educational benefit of additional cataract surgery cases in ophthalmology residency. *J Cataract Refract Surg* 2020;46(11):1495-500. doi: 10.1097/j.jcrs.0000000000000298 [published Online First: 2020/07/11]
223. Hopkins L, Robinson DBT, Brown C, et al. Trauma and Orthopedic Surgery Curriculum Concordance: An Operative Learning Curve Trajectory Perspective. *J Surg Educ* 2019;76(6):1569-78. doi: 10.1016/j.jsurg.2019.05.009 [published Online First: 2019/05/31]
224. Gupta S, Haripriya A, Vardhan SA, et al. Residents' Learning Curve for Manual Small-Incision Cataract Surgery at Aravind Eye Hospital, India. *Ophthalmology* 2018;125(11):1692-99. doi: 10.1016/j.ophtha.2018.04.033 [published Online First: 2018/06/05]
225. van Rijswijk AS, Moes DE, Geubbels N, et al. Can a laparoscopic Roux-en-Y gastric bypass be safely performed by surgical residents in a bariatric center-of-excellence? The learning curve of surgical residents in bariatric surgery. *Surg Endosc* 2018;32(2):1012-20. doi: 10.1007/s00464-017-5779-3 [published Online First: 2017/09/25]
226. Powers K, Lazarou G, Greston W, et al. The Learning Curve in Laparoscopic Reconstructive Surgery: A Pilot Study. *Female Pelvic Medicine & Reconstructive Surgery* 2005;11:329-32. doi: 10.1097/01.spv.0000193830.27759.ab
227. Umali MIN, Castillo TR. Operative Time and Complication Rates of Resident Phacoemulsification Surgeries in a National University Hospital: A Five-Year Review. *Clin Ophthalmol* 2020;14:4065-72. doi: 10.2147/opth.S283754 [published Online First: 2020/12/03]

228. Al-Jindan M, Almarshood A, Yassin SA, et al. Assessment of Learning Curve in Phacoemulsification Surgery Among the Eastern Province Ophthalmology Program Residents. *Clinical ophthalmology (Auckland, NZ)* 2020;14:113-18. doi: 10.2147/OPTH.S241250
229. Masoud M, Ibrahim A, Elemam A, et al. Learning curve of laparoscopic nephrectomy: a prospective pilot study. *African Journal of Urology* 2020;26(1):24. doi: 10.1186/s12301-020-00024-x
230. Shibuya S, Fujiwara N, Ochi T, et al. The learning curve of laparoscopic percutaneous extraperitoneal closure (LPEC) for inguinal hernia: protocoled training in a single center for six pediatric surgical trainees. *BMC Surg* 2019;19(1):6-6. doi: 10.1186/s12893-019-0470-3
231. Bracale U, Merola G, Sciuto A, et al. Achieving the Learning Curve in Laparoscopic Inguinal Hernia Repair by Tapp: A Quality Improvement Study. *J Invest Surg* 2019;32(8):738-45. doi: 10.1080/08941939.2018.1468944 [published Online First: 2018/06/15]
232. Abou-Foul AK, Taghi A, Tolley NS, et al. Assessment of progress and competence in tonsillectomy surgery using cumulative sum analysis. *J Laryngol Otol* 2018;132(8):734-38. doi: 10.1017/s0022215118001081 [published Online First: 2018/07/13]
233. Stride HP, George BC, Williams RG, et al. Relationship of procedural numbers with meaningful procedural autonomy in general surgery residents. *Surgery* 2018;163(3):488-94. doi: 10.1016/j.surg.2017.10.011 [published Online First: 2017/12/27]
234. Serrano OK, Bangdiwala AS, Vock DM, et al. Defining the Tipping Point in Surgical Performance for Laparoscopic Donor Nephrectomy Among Transplant Surgery Fellows: A Risk-Adjusted Cumulative Summation Learning Curve Analysis. *Am J Transplant* 2017;17(7):1868-78. doi: 10.1111/ajt.14187 [published Online First: 2016/12/29]
235. Joseph M, Phillips M, Farrell TM, et al. Can residents safely and efficiently be taught single incision laparoscopic cholecystectomy? *J Surg Educ* 2012;69(4):468-72. doi: 10.1016/j.jsurg.2012.03.006 [published Online First: 2012/06/09]
236. Malik MU, Varela DA, Park E, et al. Determinants of resident competence in mastoidectomy: role of interest and deliberate practice. *Laryngoscope* 2013;123(12):3162-7. doi: 10.1002/lary.24179 [published Online First: 2013/07/24]
237. Taravella MJ, Davidson R, Erlanger M, et al. Characterizing the learning curve in phacoemulsification. *J Cataract Refract Surg* 2011;37(6):1069-75. doi: 10.1016/j.jcrs.2010.12.054 [published Online First: 2011/05/21]
238. Hiemstra E, Kolkman W, Wolterbeek R, et al. Value of an objective assessment tool in the operating room. *Can J Surg* 2011;54(2):116-22. doi: 10.1503/cjs.032909
239. Jelovsek JE, Walters MD, Korn A, et al. Establishing cutoff scores on assessments of surgical skills to determine surgical competence. *Am J Obstet Gynecol* 2010;203(1):81.e1-6. doi: 10.1016/j.ajog.2010.01.073 [published Online First: 2010/04/27]
240. Maeda T, Tan KY, Konishi F, et al. Accelerated learning curve for colorectal resection, open versus laparoscopic approach, can be attained with expert supervision. *Surg Endosc* 2010;24(11):2850-4. doi: 10.1007/s00464-010-1063-5 [published Online First: 2010/05/06]

241. Francis HW, Masood H, Laeeq K, et al. Defining milestones toward competency in mastoidectomy using a skills assessment paradigm. *Laryngoscope* 2010;120(7):1417-21. doi: 10.1002/lary.20953 [published Online First: 2010/06/26]
242. Papavramidis TS, Michalopoulos N, Pliakos J, et al. Minimally invasive video-assisted total thyroidectomy: an easy to learn technique for skillful surgeons. *Head Neck* 2010;32(10):1370-6. doi: 10.1002/hed.21336 [published Online First: 2010/01/22]
243. Zhang X, Wang B, Ma X, et al. Laparoscopic adrenalectomy for beginners without open counterpart experience: initial results under staged training. *Urology* 2009;73(5):1061-5. doi: 10.1016/j.urology.2008.11.058 [published Online First: 2009/04/28]
244. Hodgins JL, Veillette C, Biau D, et al. The knee arthroscopy learning curve: quantitative assessment of surgical skills. *Arthroscopy* 2014;30(5):613-21. doi: 10.1016/j.arthro.2014.02.021 [published Online First: 2014/04/15]
245. Malik KI, Siau K, Dunckley P, et al. Colorectal trainees in the UK struggle to meet JAG certification standards in colonoscopy by the end of their training. *Colorectal Dis* 2019;21(6):715-22. doi: 10.1111/codi.14591 [published Online First: 2019/02/23]
246. Wexner SD, Litwin D, Cohen J, et al. Principles of privileging and credentialing for endoscopy and colonoscopy. *Gastrointest Endosc* 2002;55(2):145-8. doi: 10.1016/s0016-5107(02)70480-x [published Online First: 2002/01/31]
247. Shahidi N, Ou G, Telford J, et al. Establishing the learning curve for achieving competency in performing colonoscopy: a systematic review. *Gastrointestinal Endoscopy* 2014;80(3):410-16. doi: https://doi.org/10.1016/j.gie.2014.04.056
248. Sedlack RE. Training to competency in colonoscopy: assessing and defining competency standards. *Gastrointestinal Endoscopy* 2011;74(2):355-66.e2. doi: https://doi.org/10.1016/j.gie.2011.02.019
249. Seymour NE, Gallagher AG, Roman SA, et al. Analysis of errors in laparoscopic surgical procedures. *Surg Endosc* 2004;18(4):592-5. doi: 10.1007/s00464-002-8927-2 [published Online First: 2004/03/18]
250. McQueen S, McKinnon V, VanderBeek L, et al. Video-Based Assessment in Surgical Education: A Scoping Review. *Journal of Surgical Education* 2019;76(6):1645-54. doi: https://doi.org/10.1016/j.jsurg.2019.05.013
251. Kirubarajan A, Young D, Khan S, et al. Artificial Intelligence and Surgical Education: A Systematic Scoping Review of Interventions. *Journal of Surgical Education* 2022;79(2):500-15. doi: https://doi.org/10.1016/j.jsurg.2021.09.012
252. Reames BN, Bacal D, Krell RW, et al. Influence of median surgeon operative duration on adverse outcomes in bariatric surgery. *Surg Obes Relat Dis* 2015;11(1):207-13. doi: 10.1016/j.soard.2014.03.018 [published Online First: 2014/07/30]
253. Julià D, Gómez N, Codina-Cazador A. Surgical skill and complication rates after bariatric surgery. *N Engl J Med* 2014;370(3):285. doi: 10.1056/NEJMc1313890 [published Online First: 2014/01/17]
254. Darzi A, Smith S, Taffinder N. Assessing operative skill. Needs to become more objective. *BMJ (Clinical research ed)* 1999;318(7188):887-88. doi: 10.1136/bmj.318.7188.887
255. Addison P, Yoo A, Duarte-Ramos J, et al. Correlation between operative time and crowd-sourced skills assessment for robotic bariatric surgery. *Surg Endosc* 2021;35(9):5303-09. doi: 10.1007/s00464-020-08019-z [published Online First: 2020/09/25]
256. Nursal TZ, Ezer A, Belli S, et al. Reaching proficiency in laparoscopic splenectomy. *World J Gastroenterol* 2009;15(32):4005-08. doi: 10.3748/wjg.15.4005

257. Abbott KL, Chen X, Clark M, et al. Number of Operative Performance Ratings Needed to Reliably Assess the Difficulty of Surgical Procedures. *Journal of Surgical Education* 2019;76(6):e189-e92. doi: https://doi.org/10.1016/j.jsurg.2019.07.008
258. Woehr DJ, Huffcutt AI. Rater training for performance appraisal: A quantitative review. *Journal of occupational and organizational psychology* 1994;67(3):189-205.
259. Beamish AJ, Johnston MJ, Harries RL, et al. Use of the eLogbook in surgical training in the United Kingdom: A nationwide survey and consensus recommendations from the Association of Surgeons in Training. *Int J Surg* 2020;84:199-206. doi: 10.1016/j.ijsu.2020.02.045 [published Online First: 2020/03/15]
260. Hope C, Reilly JJ, Griffiths G, et al. The impact of COVID-19 on surgical training: a systematic review. *Tech Coloproctol* 2021;25(5):505-20. doi: 10.1007/s10151-020-02404-5 [published Online First: 2021/01/29]
261. Joyce DP, Ryan D, Kavanagh DO, et al. Impact of COVID-19 on operative experience of junior surgical trainees. *Br J Surg* 2021;108(1):e33-e34. doi: 10.1093/bjs/znaa079 [published Online First: 2021/03/01]
262. Scott DJ, Bergen PC, Rege RV, et al. Laparoscopic training on bench models: better and more cost effective than operating room experience? *J Am Coll Surg* 2000;191(3):272-83. doi: 10.1016/s1072-7515(00)00339-2 [published Online First: 2000/09/16]
263. Milburn JA, Khera G, Hornby ST, et al. Introduction, availability and role of simulation in surgical education and training: review of current evidence and recommendations from the Association of Surgeons in Training. *International journal of surgery (London, England)* 2012;10(8):393-8. doi: 10.1016/j.ijsu.2012.05.005 [published Online First: 2012/05/23]
264. Kanakala V, Bawa S, Gallagher P, et al. Outcome of patients in laparoscopic training courses compared to standard patients. *The surgeon : journal of the Royal Colleges of Surgeons of Edinburgh and Ireland* 2010;8(3):132-5. doi: 10.1016/j.surge.2009.10.030 [published Online First: 2010/04/20]
265. Borgersen NJ, Naur TMH, Sørensen SMD, et al. Gathering Validity Evidence for Surgical Simulation: A Systematic Review. *Annals of surgery* 2018;267(6):1063-68. doi: 10.1097/sla.0000000000002652
266. Lawaetz J, Skovbo Kristensen JS, Nayahangan LJ, et al. Simulation Based Training and Assessment in Open Vascular Surgery: A Systematic Review. *European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery* 2020 doi: 10.1016/j.ejvs.2020.11.003 [published Online First: 2020/12/15]
267. Peters MDJ GC, McInerney P, Munn Z, Tricco AC, Khalil, H. Chapter 11: Scoping Reviews (2020 version). In: Aromataris E, Munn Z (Editors). JBI Manual for Evidence Synthesis. In: Aromataris E MZ, ed. https://synthesismanual.jbi.global.: Joanna Briggs Institute, 2020.
268. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* 2018;169(7):467-73. doi: 10.7326/m18-0850 [published Online First: 2018/09/05]
269. Spencer L, Ritchie J, O'Connor W. Analysis: practices, principles and processes. *Qualitative research practice: A guide for social science students and researchers* 2003;199:218.
270. Raad WN, Ayub A, Huang CY, et al. Robotic Thoracic Surgery Training for Residency Programs: A Position Paper for an Educational Curriculum. *Innovations (Phila)*

2018;13(6):417-22. doi: 10.1097/imi.0000000000000573 [published Online First: 2018/12/06]
271. Lui JT, Compton ED, Ryu WHA, et al. Assessing the role of virtual reality training in Canadian Otolaryngology-Head & Neck Residency Programs: a national survey of program directors and residents. *J Otolaryngol Head Neck Surg* 2018;47(1):61. doi: 10.1186/s40463-018-0309-4 [published Online First: 2018/10/05]
272. Yiasemidou M, Glassman D, Tomlinson J, et al. Perceptions About the Present and Future of Surgical Simulation: A National Study of Mixed Qualitative and Quantitative Methodology. *Journal of surgical education* 2017;74(1):108-16. doi: 10.1016/j.jsurg.2016.07.011 [published Online First: 2016/09/13]
273. Aydin A, Ahmed K, Shafi AM, et al. The role of simulation in urological training - A quantitative study of practice and opinions. *The surgeon : journal of the Royal Colleges of Surgeons of Edinburgh and Ireland* 2016;14(6):301-07. doi: 10.1016/j.surge.2015.06.003 [published Online First: 2015/07/08]
274. Hagen SS, Ferguson KJ, Sharp WJ, et al. Residents' attitudes about the introduction of a surgical skills simulation laboratory. *Simul Healthc* 2010;5(1):28-32. doi: 10.1097/SIH.0b013e3181a3dfbe [published Online First: 2010/04/13]
275. Veneziano D, Poniatowski LH, Reihsen TE, et al. Preliminary evaluation of the SimPORTAL major vessel injury (MVI) repair model. *Surg Endosc* 2016;30(4):1405-12. doi: 10.1007/s00464-015-4344-1 [published Online First: 2015/07/04]
276. Ganju A, Aoun SG, Daou MR, et al. The role of simulation in neurosurgical education: a survey of 99 United States neurosurgery program directors. *World Neurosurg* 2013;80(5):e1-8. doi: 10.1016/j.wneu.2012.11.066 [published Online First: 2012/11/28]
277. Rosenthal R, Gantert WA, Hamel C, et al. The future of patient safety: Surgical trainees accept virtual reality as a new training tool. *Patient safety in surgery* 2008;2:16. doi: 10.1186/1754-9493-2-16 [published Online First: 2008/06/12]
278. Le CQ, Lightner DJ, VanderLei L, et al. The current role of medical simulation in american urological residency training programs: an assessment by program directors. *J Urol* 2007;177(1):288-91. doi: 10.1016/j.juro.2006.08.106 [published Online First: 2006/12/13]
279. Shetty S, Zevin B, Grantcharov TP, et al. Perceptions, training experiences, and preferences of surgical residents toward laparoscopic simulation training: a resident survey. *J Surg Educ* 2014;71(5):727-33. doi: 10.1016/j.jsurg.2014.01.006 [published Online First: 2014/05/06]
280. Ahmed K, Aydin A, Dasgupta P, et al. A novel cadaveric simulation program in urology. *J Surg Educ* 2015;72(4):556-65. doi: 10.1016/j.jsurg.2015.01.005 [published Online First: 2015/02/17]
281. Tranchart H, Aurégan JC, Gaillard M, et al. [Evaluation of surgical skills of French ophthalmology, orthopedic and gastrointestinal surgery residents: Current status and perspectives]. *Journal francais d'ophtalmologie* 2015;38(8):679-88. doi: 10.1016/j.jfo.2015.03.008 [published Online First: 2015/09/08]
282. Rodriguez R, Dad N, To JK. 24-Hour Simulation Lab Access and Laparoscopic Skills Improvement: A Pilot Study. *Journal of Minimally Invasive Gynecology* 2018;25(7):S52. doi: 10.1016/j.jmig.2018.09.085
283. Crochet P, Aggarwal R, Berdah S, et al. [Current and future use of surgical skills simulation in gynecologic resident education: a French national survey]. *Journal de gynecologie,*

*obstetrique et biologie de la reproduction* 2014;43(5):379-86. doi: 10.1016/j.jgyn.2013.10.003 [published Online First: 2013/12/18]
284. Kim S, Dunkin BJ, Paige JT, et al. What is the future of training in surgery? Needs assessment of national stakeholders. *Surgery* 2014;156(3):707-17. doi: 10.1016/j.surg.2014.04.047 [published Online First: 2014/09/02]
285. Scott DJ, Pugh CM, Ritter EM, et al. New directions in simulation-based surgical education and training: validation and transfer of surgical skills, use of nonsurgeons as faculty, use of simulation to screen and select surgery residents, and long-term follow-up of learners. *Surgery* 2011;149(6):735-44. doi: 10.1016/j.surg.2010.11.010 [published Online First: 2011/02/08]
286. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;79(10 Suppl):S70-81. doi: 10.1097/00001888-200410001-00022 [published Online First: 2004/09/24]
287. Gallagher AG, Leonard G, Traynor OJ. Role and feasibility of psychomotor and dexterity testing in selection for surgical training. *ANZ J Surg* 2009;79(3):108-13. doi: 10.1111/j.1445-2197.2008.04824.x [published Online First: 2009/03/26]
288. Bingmer K, Ofshteyn A, Stein SL, et al. Decline of open surgical experience for general surgery residents. *Surg Endosc* 2020;34(2):967-72. doi: 10.1007/s00464-019-06881-0 [published Online First: 2019/06/12]
289. Noureldin YA, Lee JY, McDougall EM, et al. Competency-Based Training and Simulation: Making a "Valid" Argument. *J Endourol* 2018;32(2):84-93. doi: 10.1089/end.2017.0650 [published Online First: 2018/02/14]
290. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *BJS (British Journal of Surgery)* 1997;84(2):273-78. doi: https://doi.org/10.1046/j.1365-2168.1997.02502.x
291. Lentz GM, Mandel LS, Lee D, et al. Testing surgical skills of obstetric and gynecologic residents in a bench laboratory setting: Validity and reliability. *American Journal of Obstetrics and Gynecology* 2001;184(7):1462-70. doi: https://doi.org/10.1067/mob.2001.114850
292. Fjørtoft K, Konge L, Gögenur I, et al. The Implementation Gap in Laparoscopic Simulation Training. *Scand J Surg* 2019;108(2):109-16. doi: 10.1177/1457496918798201 [published Online First: 2018/09/13]
293. Lawaetz J, Soenens G, Eiberg J, et al. Identifying a Big Implementation Gap in Simulation Based Education in Vascular Surgery in Europe: The VASSIM Study. *EJVES Vascular Forum* 2022;54:e46-e47. doi: 10.1016/j.ejvsvf.2021.12.065
294. Toale C, Morris M, Kavanagh D. Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review. *The American Journal of Surgery* 2021;222(4):723-30. doi: https://doi.org/10.1016/j.amjsurg.2021.03.035
295. Ganju A, Aoun SG, Daou MR, et al. The role of simulation in neurosurgical education: a survey of 99 United States neurosurgery program directors. *World neurosurgery* 2013;80(5):e1-e8.
296. Hagen SS, Ferguson KJ, Sharp WJ, et al. Residents' attitudes about the introduction of a surgical skills simulation laboratory. *Simulation in Healthcare* 2010;5(1):28-32.
297. Burden C, Fox R, Lenguerrand E, et al. Curriculum development for basic gynaecological laparoscopy with comparison of expert trainee opinions; prospective cross-sectional observational study. *European Journal of Obstetrics & Gynecology and Reproductive Biology* 2014;180:1-7.

298. Rosenthal R, Gantert WA, Hamel C, et al. The future of patient safety: surgical trainees accept virtual reality as a new training tool. *Patient safety in surgery* 2008;2(1):1-7.
299. Shetty S, Zevin B, Grantcharov TP, et al. Perceptions, training experiences, and preferences of surgical residents toward laparoscopic simulation training: a resident survey. *Journal of surgical education* 2014;71(5):727-33.
300. Aydin A, Ahmed K, Shafi AM, et al. The role of simulation in urological training–a quantitative study of practice and opinions. *the surgeon* 2016;14(6):301-07.
301. Veneziano D, Poniatowski LH, Reihsen TE, et al. Preliminary evaluation of the SimPORTAL major vessel injury (MVI) repair model. *Surg Endosc* 2016;30(4):1405-12.
302. Clarke V, Braun V, Hayfield N. Thematic analysis. *Qualitative psychology: A practical guide to research methods* 2015;222:248.
303. Braun V, Clarke V. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 2019;11(4):589-97.
304. Britten N. Qualitative interviews in medical research. *Bmj* 1995;311(6999):251-3. doi: 10.1136/bmj.311.6999.251 [published Online First: 1995/07/22]
305. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19(6):349-57. doi: 10.1093/intqhc/mzm042 [published Online First: 2007/09/18]
306. Kim S, Dunkin BJ, Paige JT, et al. What is the future of training in surgery? Needs assessment of national stakeholders. *Surgery* 2014;156(3):707-17. doi: https://doi.org/10.1016/j.surg.2014.04.047
307. Clarke S. Qualitative research: the essential guide to theory and practice [by] Maggi Savin-Baden & Claire Howell Major. *Psychology Learning & Teaching* 2014;13(1):66-68.
308. Toale C, Morris M, Kavanagh D. Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review. *Am J Surg* 2021 doi: 10.1016/j.amjsurg.2021.03.035 [published Online First: 2021/04/15]
309. Clarke V, Braun V, Hayfield N. Thematic analysis. *Qualitative psychology: A practical guide to research methods* 2015:222-48.
310. Sandelowski M. Sample size in qualitative research. *Res Nurs Health* 1995;18(2):179-83. doi: 10.1002/nur.4770180211 [published Online First: 1995/04/01]
311. Varpio L, Ajjawi R, Monrouxe LV, et al. Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking. *Med Educ* 2017;51(1):40-50. doi: 10.1111/medu.13124 [published Online First: 2016/12/17]
312. Dickinson KJ, Bass BL, Pei KY. Public Perceptions of General Surgery Residency Training. *Journal of Surgical Education* 2021;78(3):717-27. doi: https://doi.org/10.1016/j.jsurg.2020.09.026
313. Wu KY, Kim S, Fung K, et al. Assessing nontechnical skills in otolaryngology emergencies through simulation-based training. *Laryngoscope* 2018;128(10):2301-06. doi: 10.1002/lary.27174 [published Online First: 2018/03/25]
314. Munz Y, Moorthy K, Bann S, et al. Ceiling effect in technical skills of surgical residents. *The American Journal of Surgery* 2004;188(3):294-300. doi: https://doi.org/10.1016/j.amjsurg.2004.02.006
315. Hagelsteen K, Langegård A, Lantz A, et al. Faster acquisition of laparoscopic skills in virtual reality with haptic feedback and 3D vision. *Minim Invasive Ther Allied Technol*

2017;26(5):269-77. doi: 10.1080/13645706.2017.1305970 [published Online First: 2017/04/04]
316. Nguyen VT, Losee JE. Time- versus Competency-Based Residency Training. *Plast Reconstr Surg* 2016;138(2):527-31. doi: 10.1097/prs.0000000000002407 [published Online First: 2016/07/29]
317. Huffman EM, Choi JN, Martin JR, et al. A Competency-Based Laparoscopic Cholecystectomy Curriculum Significantly Improves General Surgery Residents' Operative Performance and Decreases Skill Variability: Cohort Study. *Ann Surg* 2021 doi: 10.1097/sla.0000000000004853 [published Online First: 2021/04/30]
318. Buckley CE, Kavanagh DO, Nugent E, et al. Zone calculation as a tool for assessing performance outcome in laparoscopic suturing. *Surg Endosc* 2015;29(6):1553-9. doi: 10.1007/s00464-014-3840-z [published Online First: 2014/10/12]
319. Munshi F, Lababidi H, Alyousef S. Low- versus high-fidelity simulations in teaching and assessing clinical skills. *Journal of Taibah University Medical Sciences* 2015;10(1):12-15. doi: https://doi.org/10.1016/j.jtumed.2015.01.008
320. Aggarwal R, Undre S, Moorthy K, et al. The simulated operating theatre: comprehensive training for surgical teams. *Qual Saf Health Care* 2004;13 Suppl 1(Suppl 1):i27-32. doi: 10.1136/qhc.13.suppl_1.i27 [published Online First: 2004/10/07]
321. Owei L, Neylan CJ, Rao R, et al. In Situ Operating Room-Based Simulation: A Review. *J Surg Educ* 2017;74(4):579-88. doi: 10.1016/j.jsurg.2017.01.001 [published Online First: 2017/03/16]
322. Gardner AK, Steffes CP, Nepomnayshy D, et al. Selection bias: Examining the feasibility, utility, and participant receptivity to incorporating simulation into the general surgery residency selection process. *The American Journal of Surgery* 2017;213(6):1171-77. doi: https://doi.org/10.1016/j.amjsurg.2016.09.029
323. Toale C, Morris M, Ryan DM, et al. Baseline Assessments of Psychomotor Aptitude are Associated with the Future Operative Performance of Surgical Trainees A Prospective Observational Cohort Study. *Annals of Surgery* 9900
324. Carey JM, Rossler K. The How When Why of High Fidelity Simulation. StatPearls. Treasure Island (FL): StatPearls Publishing
Copyright © 2022, StatPearls Publishing LLC. 2022.
325. Sturm LP, Windsor JA, Cosman PH, et al. A systematic review of skills transfer after surgical simulation training. *Ann Surg* 2008;248(2):166-79. doi: 10.1097/SLA.0b013e318176bf24 [published Online First: 2008/07/25]
326. Gallagher AG, Neary P, Gillen P, et al. Novel method for assessment and selection of trainees for higher surgical training in general surgery. *ANZ Journal of Surgery* 2008;78(4):282-90. doi: 10.1111/j.1445-2197.2008.04439.x
327. Artino AR, Jr., La Rochelle JS, Dezee KJ, et al. Developing questionnaires for educational research: AMEE Guide No. 87. *Medical teacher* 2014;36(6):463-74. doi: 10.3109/0142159X.2014.889814 [published Online First: 2014/03/24]
328. Toale C, Morris M, Kavanagh D. Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review. *The American Journal of Surgery* 2021 doi: https://doi.org/10.1016/j.amjsurg.2021.03.035
329. Toale C, Morris M, Kavanagh DO. Perspectives on simulation-based assessment of operative skill in surgical training. *Medical Teacher* 2023;45(3):313-20. doi: 10.1080/0142159X.2022.2134001

330. Willis GB, Artino AR, Jr. What Do Our Respondents Think We're Asking? Using Cognitive Interviewing to Improve Medical Education Surveys. *Journal of graduate medical education* 2013;5(3):353-56. doi: 10.4300/JGME-D-13-00154.1
331. Lei C. A Psychometric Evaluation of 4-Point and 6-Point Likert-Type Scales in Relation to Reliability and Validity. *Applied Psychological Measurement* 1994;18(3):205-15. doi: 10.1177/014662169401800302
332. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006;3:77-101. doi: 10.1191/1478088706qp063oa
333. Aydin A, Ahmed K, Shafi AMA, et al. The role of simulation in urological training – A quantitative study of practice and opinions. *The Surgeon* 2016;14(6):301-07. doi: https://doi.org/10.1016/j.surge.2015.06.003
334. Toft LEB, Bottinor W, Cobourn A, et al. A simulation-enhanced, spaced learning, interprofessional "code blue" curriculum improves ACLS algorithm adherence and trainee resuscitation skill confidence. *J Interprof Care* 2023;37(4):623-28. doi: 10.1080/13561820.2022.2140130 [published Online First: 2022/11/15]
335. Gardner AK, Smink DS, Scott BG, et al. How Much Are We Spending on Resident Selection? *Journal of surgical education* 2018;75(6):e85-e90. doi: https://doi.org/10.1016/j.jsurg.2018.10.001
336. Solomou G, Murphy S, Bandyopadhyay S, et al. Neurosurgery specialty training in the UK: What you need to know to be shortlisted for an interview. *Annals of Medicine and Surgery* 2020;57:287-90. doi: https://doi.org/10.1016/j.amsu.2020.07.047
337. Filiberto AC, Cooper LA, Loftus TJ, et al. Objective predictors of intern performance. *BMC Medical Education* 2021;21(1):77. doi: 10.1186/s12909-021-02487-0
338. Lammy S. Core Surgical Training Applications: Making The Cut. *Bulletin of The Royal College of Surgeons of England* 2014;95:48-49. doi: 10.1308/147363513X13690603817823
339. Marwan Y, Waly F, Algarni N, et al. The Role of Letters of Recommendation in the Selection Process of Surgical Residents in Canada: A National Survey of Program Directors. *J Surg Educ* 2017;74(4):762-67. doi: 10.1016/j.jsurg.2017.01.006 [published Online First: 2017/01/28]
340. Patel NM, Khajuria A, Khajuria A. Utility of a webinar to educate trainees on UK core surgical training (CST) selection–A cross sectional study and future implications amidst the COVID-19 pandemic. *Annals of Medicine and Surgery* 2020;59:35-40.
341. Gardner AK, Grantcharov T, Dunkin BJ. The Science of Selection: Using Best Practices From Industry to Improve Success in Surgery Training. *Journal of Surgical Education* 2018;75(2):278-85. doi: https://doi.org/10.1016/j.jsurg.2017.07.010
342. DeVaul RA, Jervey F, Chappell JA, et al. Medical School Performance of Initially Rejected Students. *JAMA* 1987;257(1):47-51. doi: 10.1001/jama.1987.03390010051027
343. Milstein RM, Wilkinson L, Burrow GN, et al. Admission decisions and performance during medical school. *Academic Medicine* 1981;56(2):77-82.
344. Kleshinski J, Khuder SA, Shapiro JI, et al. Impact of preadmission variables on USMLE step 1 and step 2 performance. *Adv Health Sci Educ Theory Pract* 2009;14(1):69-78. doi: 10.1007/s10459-007-9087-x [published Online First: 2007/11/08]
345. Gardner AK, D'Onofrio BC, Dunkin BJ. Can We Get Faculty Interviewers on the Same Page? An Examination of a Structured Interview Course for Surgeons. *Journal of Surgical Education* 2018;75(1):72-77. doi: https://doi.org/10.1016/j.jsurg.2017.06.006

346. Gardner AK, Dunkin BJ. Evaluation of validity evidence for personality, emotional intelligence, and situational judgment tests to identify successful residents. *JAMA surgery* 2018;153(5):409-16.
347. Ireland RCoSi. ST1 Interviews 2021 - Selection Process & Marking Guide, Core Surgical Training: Royal College of Surgeons in Ireland 2020 [updated October 2020; cited 2021 11/06/2021]. Available from: https://www.rcsi.com/surgery/training/surgery/cst/apply accessed 11/06/21 2021.
348. Panait L, Larios JM, Brenes RA, et al. Surgical Skills Assessment of Applicants to General Surgery Residency. *Journal of Surgical Research* 2011;170(2):189-94. doi: https://doi.org/10.1016/j.jss.2011.04.006
349. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;370(9596):1453-7. doi: 10.1016/s0140-6736(07)61602-x [published Online First: 2007/12/08]
350. Henn P, Gallagher AG, Nugent E, et al. A computerised test of perceptual ability for learning endoscopic and laparoscopic surgery and other image guided procedures: Score norms for PicSOr. *Am J Surg* 2017;214(5):969-73. doi: 10.1016/j.amjsurg.2017.01.025 [published Online First: 2017/02/24]
351. Gallagher AG, Cowie R, Crothers I, et al. PicSOr: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surg Endosc* 2003;17(9):1468-71. doi: 10.1007/s00464-002-8569-4 [published Online First: 2003/06/13]
352. Cowie R. Measurement and modelling of perceived slant in surfaces represented by freely viewed line drawings. *Perception* 1998;27(5):505-40. doi: 10.1068/p270505 [published Online First: 1999/03/10]
353. Kolozsvari NO, Andalib A, Kaneva P, et al. Sex is not everything: the role of gender in early performance of a fundamental laparoscopic skill. *Surg Endosc* 2011;25(4):1037-42. doi: 10.1007/s00464-010-1311-8 [published Online First: 2010/08/25]
354. Haluck RS, Gallagher AG, Satava RM, et al. Reliability and validity of Endotower, a virtual reality trainer for angled endoscope navigation. *Stud Health Technol Inform* 2002;85:179-84. [published Online First: 2004/10/02]
355. Ritter EM, McClusky DA, 3rd, Gallagher AG, et al. Perceptual, visuospatial, and psychomotor abilities correlate with duration of training required on a virtual-reality flexible endoscopy simulator. *Am J Surg* 2006;192(3):379-84. doi: 10.1016/j.amjsurg.2006.03.003 [published Online First: 2006/08/22]
356. Westman B, Ritter EM, Kjellin A, et al. Visuospatial abilities correlate with performance of senior endoscopy specialist in simulated colonoscopy. *J Gastrointest Surg* 2006;10(4):593-9. doi: 10.1016/j.gassur.2005.08.014 [published Online First: 2006/04/22]
357. Arora H, Uribe J, Ralph W, et al. Assessment of construct validity of the endoscopic sinus surgery simulator. *Arch Otolaryngol Head Neck Surg* 2005;131(3):217-21. doi: 10.1001/archotol.131.3.217 [published Online First: 2005/03/23]
358. Ekstrom RB, Harman HH. Manual for kit of factor-referenced cognitive tests, 1976: Educational testing service 1976.
359. Carroll JB. Human cognitive abilities: A survey of factor-analytic studies: Cambridge University Press 1993.

360. Manual for kit of factor-referenced cognitive tests, 1976. In: Ekstrom RB, Harman HH, eds. Princeton, N.J. :: Educational Testing Service, 1976.
361. Dikmen SS, Heaton RK, Grant I, et al. Test–retest reliability and practice effects of Expanded Halstead–Reitan Neuropsychological Test Battery. *Journal of the International Neuropsychological Society* 1999;5(4):346-56. doi: 10.1017/S1355617799544056 [published Online First: 1999/05/01]
362. Harrington C, Kavanagh D, Ryan D, et al. Objective scoring of electronic surgical logbooks: Analysis of impact and observations within a surgical training body. *The American Journal of Surgery* 2017;214 doi: 10.1016/j.amjsurg.2017.07.028
363. Bujang MA, Baharum N. Sample Size Guideline for Correlation Analysis. *World Journal of Social Science Research* 2016;3:37. doi: 10.22158/wjssr.v3n1p37
364. Myers RH, Myers RH. Classical and modern regression with applications: Duxbury press Belmont, CA 1990.
365. Cohen J. Statistical power analysis for the behavioral sciences: Academic press 2013.
366. Buckley CE, Kavanagh DO, Traynor O, et al. Does aptitude matter? *Journal of the American College of Surgeons* 2012;215(3, Supplement):S116. doi: https://doi.org/10.1016/j.jamcollsurg.2012.06.304
367. Ritter EM, McClusky DA, Gallagher AG, et al. Perceptual, visuospatial, and psychomotor abilities correlate with duration of training required on a virtual-reality flexible endoscopy simulator. *The American Journal of Surgery* 2006;192(3):379-84. doi: https://doi.org/10.1016/j.amjsurg.2006.03.003
368. Wanzel KR, Hamstra SJ, Anastakis DJ, et al. Effect of visual-spatial ability on learning of spatially-complex surgical skills. *The lancet* 2002;359(9302):230-31.
369. Wanzel KR, Hamstra SJ, Caminiti MF, et al. Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. *Surgery* 2003;134(5):750-57. doi: https://doi.org/10.1016/S0039-6060(03)00248-4
370. Nugent E, Hseino H, Boyle E, et al. Assessment of the role of aptitude in the acquisition of advanced laparoscopic surgical skill sets. *International Journal of Colorectal Disease* 2012;27(9):1207-14. doi: 10.1007/s00384-012-1458-y
371. Henn P, Gallagher AG, Nugent E, et al. Visual spatial ability for surgical trainees: implications for learning endoscopic, laparoscopic surgery and other image-guided procedures. *Surgical endoscopy* 2018;32(8):3634-39. doi: 10.1007/s00464-018-6094-3
372. Stefanidis D, Korndorffer JR, Jr., Black FW, et al. Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. *Surgery* 2006;140(2):252-62. doi: 10.1016/j.surg.2006.04.002 [published Online First: 2006/08/15]
373. The general surgery milestone project. *J Grad Med Educ* 2014;6(1 Suppl 1):320-8. doi: 10.4300/JGME-06-01s1-40.1 [published Online First: 2014/04/05]
374. Lund J. The new General Surgical Curriculum and ISCP. *Surgery (Oxf)* 2020;38(10):601-06. doi: 10.1016/j.mpsur.2020.07.005 [published Online First: 2020/09/10]
375. McCulloch P, Mishra A, Handa A, et al. The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre. *Qual Saf Health Care* 2009;18(2):109-15. doi: 10.1136/qshc.2008.032045 [published Online First: 2009/04/04]
376. Abd Elwahab SM, Cosgrave J, Doherty E, et al. Grit personality trait of doctors and medical students in level 4 acute hospital. *Surgeon* 2021 doi: 10.1016/j.surge.2021.04.010 [published Online First: 2021/07/05]

377. Abi-Jaoudé JG, Kennedy-Metz LR, Dias RD, et al. Measuring and Improving Emotional Intelligence in Surgery: A Systematic Review. *Ann Surg* 2022;275(2):e353-e60. doi: 10.1097/sla.0000000000005022 [published Online First: 2021/06/26]
378. Hagelsteen K, Pedersen H, Bergenfelz A, et al. Different approaches to selection of surgical trainees in the European Union. *BMC Medical Education* 2021;21(1):363. doi: 10.1186/s12909-021-02779-5
379. LaGrasso JR, Kennedy DA, Hoehn JG, et al. Selection criteria for the integrated model of plastic surgery residency. *Plast Reconstr Surg* 2008;121(3):121e-25e. doi: 10.1097/01.prs.0000299456.96822.1b [published Online First: 2008/03/05]
380. Makdisi G, Takeuchi T, Rodriguez J, et al. How we select our residents--a survey of selection criteria in general surgery residents. *J Surg Educ* 2011;68(1):67-72. doi: 10.1016/j.jsurg.2010.10.003 [published Online First: 2011/02/05]
381. Janis JE, Hatef DA. Resident selection protocols in plastic surgery: a national survey of plastic surgery program directors. *Plast Reconstr Surg* 2008;122(6):1929-39. doi: 10.1097/PRS.0b013e31818d20ae [published Online First: 2008/12/04]
382. Makdisi G, Takeuchi T, Rodriguez J, et al. How We Select Our Residents—A Survey of Selection Criteria in General Surgery Residents. *Journal of Surgical Education* 2011;68(1):67-72. doi: https://doi.org/10.1016/j.jsurg.2010.10.003
383. Selber JC, Tong W, Koshy J, et al. Correlation Between Trainee Candidate Selection Criteria and Subsequent Performance. *Journal of the American College of Surgeons* 2014;219(5):951-57. doi: https://doi.org/10.1016/j.jamcollsurg.2014.07.942
384. Roberts C, Khanna P, Rigby L, et al. Utility of selection methods for specialist medical training: A BEME (best evidence medical education) systematic review: BEME guide no. 45. *Med Teach* 2018;40(1):3-19. doi: 10.1080/0142159x.2017.1367375 [published Online First: 2017/08/30]
385. Patterson F, Ferguson E, Zibarras L. Selection into Medical Education and Training. Understanding Medical Education2018:375-88.
386. Longo WE, Seashore J, Duffy A, et al. Attrition of categoric general surgery residents: results of a 20-year audit. *Am J Surg* 2009;197(6):774-8; discussion 79-80. doi: 10.1016/j.amjsurg.2008.06.038 [published Online First: 2009/01/31]
387. Yaghoubian A, Galante J, Kaji A, et al. General Surgery Resident Remediation and Attrition: A Multi-institutional Study. *Archives of Surgery* 2012;147(9):829-33. doi: 10.1001/archsurg.2012.1676
388. Torbeck L, Canal DF. Remediation practices for surgery residents. *The American Journal of Surgery* 2009;197(3):397-402. doi: https://doi.org/10.1016/j.amjsurg.2008.10.008
389. Eva KW, Reiter HI, Trinh K, et al. Predictive validity of the multiple mini-interview for selecting medical trainees. *Med Educ* 2009;43(8):767-75. doi: 10.1111/j.1365-2923.2009.03407.x [published Online First: 2009/08/08]
390. Patterson F, Knight A, Dowell J, et al. How effective are selection methods in medical education? A systematic review. *Med Educ* 2016;50(1):36-60. doi: 10.1111/medu.12817 [published Online First: 2015/12/24]
391. Tolan AM, Kaji AH, Quach C, et al. The electronic residency application service application can predict accreditation council for graduate medical education competency-based surgical resident performance. *J Surg Educ* 2010;67(6):444-8. doi: 10.1016/j.jsurg.2010.05.002 [published Online First: 2010/12/16]

392. Roberts C, Khanna P, Rigby L, et al. Utility of selection methods for specialist medical training: a BEME (best evidence medical education) systematic review: BEME guide no. 45. *Medical teacher* 2018;40(1):3-19.
393. Kenny S, McInnes M, Singh V. Associations between residency selection strategies and doctor performance: a meta-analysis. *Med Educ* 2013;47(8):790-800. doi: 10.1111/medu.12234 [published Online First: 2013/07/11]
394. Panda N, Bahdila D, Abdullah A, et al. Association Between USMLE Step 1 Scores and In-Training Examination Performance: A Meta-Analysis. *Acad Med* 2021;96(12):1742-54. doi: 10.1097/acm.0000000000004227 [published Online First: 2021/07/30]
395. Scrimgeour DSG, Cleland J, Lee AJ, et al. Impact of performance in a mandatory postgraduate surgical examination on selection into specialty training. *BJS Open* 2017;1(3):67-74. doi: 10.1002/bjs5.7 [published Online First: 2018/06/29]
396. Louridas M, Szasz P, de Montbrun S, et al. Can we predict technical aptitude? *Annals of surgery* 2016;263(4):673-91.
397. Schaverien MV. Selection for surgical training: an evidence-based review. *Journal of surgical education* 2016;73(4):721-29.
398. Louridas M, Szasz P, de Montbrun S, et al. Optimizing the selection of general surgery residents: a national consensus. *Journal of surgical education* 2017;74(1):100-07.
399. Toale C, Morris M, Ryan DM, et al. Baseline Assessments of Psychomotor Aptitude are Associated with the Future Operative Performance of Surgical Trainees A Prospective Observational Cohort Study. *Annals of Surgery* 9900:10.1097/SLA.0000000000005565. doi: 10.1097/sla.0000000000005565
400. Iversen ED, Wolderslund MO, Kofoed P-E, et al. Codebook for rating clinical communication skills based on the Calgary-Cambridge Guide. *BMC Medical Education* 2020;20(1):140. doi: 10.1186/s12909-020-02050-3
401. Harrington CM, Kavanagh DO, Ryan D, et al. Objective scoring of an electronic surgical logbook: Analysis of impact and observations within a surgical training body. *Am J Surg* 2017;214(5):962-68. doi: 10.1016/j.amjsurg.2017.07.028 [published Online First: 2017/08/07]
402. Newble D, Dawson B, Dauphinee D, et al. Guidelines for assessing clinical competence. *Teaching and Learning in Medicine: An International Journal* 1994;6(3):213-20.
403. Nally DM, Kavanagh D, Traynor O, et al. Attrition From a National Surgical Training Program: A Qualitative Analysis. *Ann Surg* 2022;275(4):621-28. doi: 10.1097/sla.0000000000004852 [published Online First: 2021/04/30]
404. Sutton PA, Beamish AJ, Rashid S, et al. Attributes of excellent surgical trainers: An analysis of outstanding trainers. *International Journal of Surgery* 2018;52:371-75. doi: https://doi.org/10.1016/j.ijsu.2017.10.007
405. Roff S, McAleer S. What is educational climate? *Med Teach* 2001;23(4):333-34. doi: 10.1080/01421590120063312 [published Online First: 2002/07/06]
406. Hashimoto DA, Sirimanna P, Gomez ED, et al. Deliberate practice enhances quality of laparoscopic surgical performance in a randomized controlled trial: from arrested development to expert performance. *Surg Endosc* 2015;29(11):3154-62. doi: 10.1007/s00464-014-4042-4 [published Online First: 2014/12/30]
407. Burgess A, Roberts C, Clark T, et al. The social validity of a national assessment centre for selection into general practice training. *BMC Medical Education* 2014;14(1):261. doi: 10.1186/s12909-014-0261-6

408. Van Der Vleuten CPM. The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education* 1996;1(1):41-67. doi: 10.1007/BF00596229
409. Abd Elwahab SM, Cosgrave J, Doherty E, et al. Grit personality trait of doctors and medical students in level 4 acute hospital. *Surgeon* 2022;20(4):216-24. doi: 10.1016/j.surge.2021.04.010 [published Online First: 2021/07/05]
410. Ellis R, Brennan PA, Lee AJ, et al. Differential attainment at MRCS according to gender, ethnicity, age and socioeconomic factors: a retrospective cohort study. *J R Soc Med* 2022;115(7):257-72. doi: 10.1177/01410768221079018 [published Online First: 2022/02/17]
411. Hurreiz H. The evolution of surgical training in the UK. *Adv Med Educ Pract* 2019;10:163-68. doi: 10.2147/amep.S189298 [published Online First: 2019/04/18]
412. Crebbin W, Watters D. Defining Standards for a Competency-Based Surgical Training Program. *Journal of Graduate Medical Education* 2012;4(4):555-55. doi: 10.4300/1949-8357-4.4.555
413. Abdelaal A. Procedure-based assessments: the past, the present and the future. *Postgrad Med J* 2020;96(1131):7-8. doi: 10.1136/postgradmedj-2019-136794 [published Online First: 2019/07/28]
414. van de Ridder JM, Stokking KM, McGaghie WC, et al. What is feedback in clinical education? *Med Educ* 2008;42(2):189-97. doi: 10.1111/j.1365-2923.2007.02973.x [published Online First: 2008/01/31]
415. What is the difference between formative and summative assessment? cmu.edu: Carnegie Mellon University [Available from: https://www.cmu.edu/teaching/assessment/basics/formative-summative.html accessed 22/04/2021 2021.
416. Bing-You RG, Trowbridge RL. Why medical educators may be failing at feedback. *Jama* 2009;302(12):1330-1. doi: 10.1001/jama.2009.1393 [published Online First: 2009/09/24]
417. Kluger AN, Van Dijk D. Feedback, the various tasks of the doctor, and the feedforward alternative. *Med Educ* 2010;44(12):1166-74. doi: 10.1111/j.1365-2923.2010.03849.x [published Online First: 2010/11/26]
418. Archer JC. State of the science in health professional education: effective feedback. *Med Educ* 2010;44(1):101-8. doi: 10.1111/j.1365-2923.2009.03546.x [published Online First: 2010/01/19]
419. Beamish AJ, Johnston MJ, Harries RL, et al. Real-world use of workplace based assessments in surgical training: A UK nationwide cross-sectional exploration of trainee perspectives and consensus recommendations from the Association of Surgeons in Training. *Int J Surg* 2020;84:212-18. doi: 10.1016/j.ijsu.2020.07.068 [published Online First: 2020/09/09]
420. Gauthier S, Cavalcanti R, Goguen J, et al. Deliberate practice as a framework for evaluating feedback in residency training. *Medical Teacher* 2015;37(6):551-57. doi: 10.3109/0142159X.2014.956059
421. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med* 2008;15(11):988-94. doi: 10.1111/j.1553-2712.2008.00227.x [published Online First: 2008/09/10]

422. Gauthier S, Cavalcanti R, Goguen J, et al. Deliberate practice as a framework for evaluating feedback in residency training. *Med Teach* 2015;37(6):551-7. doi: 10.3109/0142159x.2014.956059 [published Online First: 2014/12/17]
423. Cantillon P, Sargeant J. Giving feedback in clinical settings. *BMJ* 2008;337:a1961. doi: 10.1136/bmj.a1961
424. Prashanti E, Ramnarayan K. Ten maxims of formative assessment. *Advances in Physiology Education* 2019;43(2):99-102. doi: 10.1152/advan.00173.2018
425. Beamish AJ, Johnston MJ, Harries RL, et al. Real-world use of workplace based assessments in surgical training: A UK nationwide cross-sectional exploration of trainee perspectives and consensus recommendations from the Association of Surgeons in Training. *International Journal of Surgery* 2020;84:212-18. doi: https://doi.org/10.1016/j.ijsu.2020.07.068
426. Abdelsattar JM, AlJamal YN, Ruparel RK, et al. Correlation of Objective Assessment Data With General Surgery Resident In-Training Evaluation Reports and Operative Volumes. *Journal of surgical education* 2018;75(6):1430-36. doi: 10.1016/j.jsurg.2018.04.016 [published Online First: 2018/05/19]
427. Yepes-Rios M, Dudek N, Duboyce R, et al. The failure to fail underperforming trainees in health professions education: A BEME systematic review: BEME Guide No. 42. *Med Teach* 2016;38(11):1092-99. doi: 10.1080/0142159x.2016.1215414 [published Online First: 2016/11/03]
428. Kozin ED, Bohnen JD, George BC, et al. Novel Mobile App Allows for Fast and Validated Intraoperative Assessment of Otolaryngology Residents. *OTO Open* 2017;1(1):2473974X16685705-2473974X05. doi: 10.1177/2473974X16685705
429. Cogbill TH, Malangoni MA, Potts JR, et al. The general surgery milestones project. *J Am Coll Surg* 2014;218(5):1056-62. doi: 10.1016/j.jamcollsurg.2014.02.016 [published Online First: 2014/04/01]
430. Toale C, Morris M, Kavanagh DO. Assessing Operative Skill in the Competency-based Education Era: Lessons From the UK and Ireland. *Ann Surg* 2022;275(4):e615-e25. doi: 10.1097/sla.0000000000005242 [published Online First: 2022/02/08]
431. Gauvin G, Hay K, Hopman W, et al. Competency-based education in general surgery: Are Canadian residents ready? *Can J Surg* 2021;64(5):E473-e75. doi: 10.1503/cjs.011520 [published Online First: 2021/09/02]
432. Wijnhoven BPL. Surgical training and research in the Netherlands. *Keio J Med* 2019;68(3):68. doi: 10.2302/kjm.68-005-ABST [published Online First: 2019/09/27]
433. Frank JR, Mungroo R, Ahmad Y, et al. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Medical Teacher* 2010;32(8):631-37. doi: 10.3109/0142159X.2010.500898
434. Holmboe ES, Sherbino J, Long DM, et al. The role of assessment in competency-based medical education. *Medical teacher* 2010;32(8):676-82.
435. Crossley J, Davies H, Humphris G, et al. Generalisability: a key to unlock professional assessment. *Med Educ* 2002;36(10):972-8. doi: 10.1046/j.1365-2923.2002.01320.x [published Online First: 2002/10/23]
436. Williams RG, Verhulst S, Colliver JA, et al. Assuring the reliability of resident performance appraisals: More items or more observations? *Surgery* 2005;137(2):141-47. doi: https://doi.org/10.1016/j.surg.2004.06.011
437. Brecknell JL, P. Core Surgical Training Curriculum iscp.ac.uk: Intercollegiate Surgical Curriculum Programme 2021 [cited 2021 28/04/2021]. Available from:

https://www.iscp.ac.uk/media/1111/core-surgical-training-curriculum-aug-2021-approved-oct-20.pdf accessed 28/04/2021 2021.
438. Crossley J, Davies H, Humphris G, et al. Generalisability: a key to unlock professional assessment. *Medical Education* 2002;36(10):972-78. doi: https://doi.org/10.1046/j.1365-2923.2002.01320.x
439. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use: Oxford University Press, USA 2015.
440. Atilgan H. Sample Size for Estimation of G and Phi Coefficients in Generalizability Theory. *Egitim Arastirmalari - Eurasian Journal of Educational Research* 2013;13:215-28.
441. Williams RG, Swanson DB, Fryer JP, et al. How Many Observations are Needed to Assess a Surgical Trainee's State of Operative Competency? *Ann Surg* 2019;269(2):377-82. doi: 10.1097/sla.0000000000002554 [published Online First: 2017/10/25]
442. Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ* 2004;38(9):1006-12. doi: 10.1111/j.1365-2929.2004.01932.x [published Online First: 2004/08/26]
443. Hodges B, Regehr G, McNaughton N, et al. OSCE checklists do not capture increasing levels of expertise. *Acad Med* 1999;74(10):1129-34. doi: 10.1097/00001888-199910000-00017 [published Online First: 1999/10/28]
444. Regehr G, MacRae H, Reznick RK, et al. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998;73(9):993-7. doi: 10.1097/00001888-199809000-00020 [published Online First: 1998/10/06]
445. Ringsted C, Østergaard D, Ravn L, et al. A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills. *Med Teach* 2003;25(6):654-8. doi: 10.1080/0142159031000160564 2 [published Online First: 2004/09/17]
446. Ilgen JS, Ma IW, Hatala R, et al. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ* 2015;49(2):161-73. doi: 10.1111/medu.12621 [published Online First: 2015/01/30]
447. Williams RG, Verhulst S, Mellinger JD, et al. Is a Single-Item Operative Performance Rating Sufficient? *Journal of Surgical Education* 2015;72(6):e212-e17. doi: https://doi.org/10.1016/j.jsurg.2015.05.002
448. Ott MC, Pack R, Cristancho S, et al. "The Most Crushing Thing": Understanding Resident Assessment Burden in a Competency-Based Curriculum. *Journal of Graduate Medical Education* 2022;14(5):583-92. doi: 10.4300/jgme-d-22-00050.1
449. Williams RG, Sanfey H, Chen XP, et al. A controlled study to determine measurement conditions necessary for a reliable and valid operative performance assessment: a controlled prospective observational study. *Ann Surg* 2012;256(1):177-87. doi: 10.1097/SLA.0b013e31825b6de4 [published Online First: 2012/07/04]
450. Goldenberg MG, Grantcharov TP. Video-analysis for the assessment of practical skill. *Tijdschrift voor Urologie* 2016;6(8):128-36. doi: 10.1007/s13629-016-0156-x
451. Lavanchy JL, Zindel J, Kirtac K, et al. Automation of surgical skill assessment using a three-stage machine learning algorithm. *Scientific Reports* 2021;11(1):5197. doi: 10.1038/s41598-021-84295-6
452. Williams RG, George BC, Bohnen JD, et al. A Proposed Blueprint for Operative Performance Training, Assessment, and Certification. *Ann Surg* 2021;273(4):701-08. doi: 10.1097/sla.0000000000004467 [published Online First: 2020/11/18]

453. Erfani Khanghahi M, Ebadi Fard Azar F. Direct observation of procedural skills (DOPS) evaluation method: Systematic review of evidence. *Med J Islam Repub Iran* 2018;32:45. doi: 10.14196/mjiri.32.45 [published Online First: 2018/08/31]
454. Bohnen JD, George BC, Williams RG, et al. The Feasibility of Real-Time Intraoperative Performance Assessment With SIMPL (System for Improving and Measuring Procedural Learning): Early Experience From a Multi-institutional Trial. *Journal of Surgical Education* 2016;73(6):e118-e30. doi: https://doi.org/10.1016/j.jsurg.2016.08.010
455. Elfenbein DM. Confidence Crisis Among General Surgery Residents: A Systematic Review and Qualitative Discourse Analysis. *JAMA Surg* 2016;151(12):1166-75. doi: 10.1001/jamasurg.2016.2792 [published Online First: 2016/09/08]
456. Lindeman B, Sarosi GA. Competency-based resident education: The United States perspective. *Surgery* 2020;167(5):777-81. doi: https://doi.org/10.1016/j.surg.2019.05.059
457. Gagnon JC, Barber BR. The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation. The SAGE encyclopedia of educational research, measurement and evaluation: Sage 2018:668.
458. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Advances in Simulation* 2016;1(1):1-12.
459. Cook DA, Lineberry M. Consequences Validity Evidence: Evaluating the Impact of Educational Assessments. *Acad Med* 2016;91(6):785-95. doi: 10.1097/acm.0000000000001114 [published Online First: 2016/02/04]
460. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;84(2):273-8. doi: 10.1046/j.1365-2168.1997.02502.x [published Online First: 1997/02/01]
461. Brown C, Ross S, Cleland J, et al. Money makes the (medical assessment) world go round: The cost of components of a summative final year Objective Structured Clinical Examination (OSCE). *Med Teach* 2015;37(7):653-59. doi: 10.3109/0142159x.2015.1033389 [published Online First: 2015/04/30]
462. Hosny SG, Johnston MJ, Pucher PH, et al. Barriers to the implementation and uptake of simulation-based training programs in general surgery: a multinational qualitative study. *J Surg Res* 2017;220:419-26.e2. doi: 10.1016/j.jss.2017.07.020 [published Online First: 2017/08/29]
463. Toale C, Morris M, Kavanagh DO. Training and assessment using the LapSim laparoscopic simulator: a scoping review of validity evidence. *Surg Endosc* 2022 doi: 10.1007/s00464-022-09593-0 [published Online First: 2022/09/20]
464. Carroll SM, Kennedy AM, Traynor O, et al. Objective assessment of surgical performance and its impact on a national selection programme of candidates for higher surgical training in plastic surgery. *Journal of Plastic, Reconstructive & Aesthetic Surgery* 2009;62(12):1543-49. doi: https://doi.org/10.1016/j.bjps.2008.06.054
465. Lord FM, Novick MR, Birnbaum A. Statistical theories of mental test scores. Oxford, England: Addison-Wesley 1968.
466. van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach* 2012;34(3):205-14. doi: 10.3109/0142159x.2012.652239 [published Online First: 2012/03/01]
467. Wass V, Van der Vleuten C, Shatzer J, et al. Assessment of clinical competence. *The Lancet* 2001;357(9260):945-49. doi: https://doi.org/10.1016/S0140-6736(00)04221-5

468. Swanson DB, van der Vleuten CP. Assessment of clinical skills with standardized patients: state of the art revisited. *Teach Learn Med* 2013;25 Suppl 1:S17-25. doi: 10.1080/10401334.2013.842916 [published Online First: 2013/11/28]
469. Niitsu H, Hirabayashi N, Yoshimitsu M, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surg Today* 2013;43(3):271-75. doi: 10.1007/s00595-012-0313-7 [published Online First: 2012/09/01]
470. Plake BS, Cizek GJ, Cizek G. The modified Angoff, extended Angoff, and Yes/No standard setting methods. *Setting performance standards Foundations, methods, and innovations* 2012:181-253.
471. Moonen–van Loon JMW, Overeem K, Govaerts MJB, et al. The Reliability of Multisource Feedback in Competency-Based Assessment Programs: The Effects of Multiple Occasions and Assessor Groups. *Academic Medicine* 2015;90(8):1093-99. doi: 10.1097/acm.0000000000000763
472. Gruppen LD, Davis WK, Fitzgerald JT, et al. Reliability, Number of Stations, and Examination Length in an Objective Structured Clinical Examination. In: Scherpbier AJJA, van der Vleuten CPM, Rethans JJ, et al., eds. Advances in Medical Education. Dordrecht: Springer Netherlands 1997:441-42.
473. Gormley G. Summative OSCEs in undergraduate medical education. *Ulster Med J* 2011;80(3):127-32.
474. Crossley J, Marriott J, Purdie H, et al. Prospective observational study to evaluate NOTSS (Non-Technical Skills for Surgeons) for assessing trainees' non-technical performance in the operating theatre. *Br J Surg* 2011;98(7):1010-20. doi: 10.1002/bjs.7478 [published Online First: 2011/04/12]
475. Toale C, Nally DM, Ryan DM, et al. How do Trainers and Trainees use Formative Workplace-based Assessments of Operative Competence to Deliver and Receive Structured Feedback? *Journal of Surgical Education* 2022;79(2):485-91. doi: https://doi.org/10.1016/j.jsurg.2021.08.023
476. Frank JR, Langer B. Collaboration, Communication, Management, and Advocacy: Teaching Surgeons New Skills through the CanMEDS Project. *World Journal of Surgery* 2003;27(8):972-78. doi: 10.1007/s00268-003-7102-9
477. Cogbill TH, Swing SR. Development of the educational milestones for surgery. *J Grad Med Educ* 2014;6(1 Suppl 1):317-9. doi: 10.4300/JGME-06-01s1-39 [published Online First: 2014/04/05]
478. Toale C, O'Byrne A, Morris M, et al. Defining operative experience targets in surgical training: A systematic review. *Surgery* 2022 doi: https://doi.org/10.1016/j.surg.2022.07.015
479. Fink A, Kosecoff J, Chassin M, et al. Consensus methods: characteristics and guidelines for use. *American journal of public health* 1984;74(9):979-83. doi: 10.2105/ajph.74.9.979
480. Dalkey NC. The Delphi Method: An Experimental Study of Group Opinion. Santa Monica, CA: RAND Corporation 1969.
481. Hsu C-C, Sandford B. The Delphi Technique: Making Sense Of Consensus. *Practical Assessment, Research and Evaluation* 2007;12
482. Nayahangan LJ, Clementsen PF, Paltved C, et al. Identifying Technical Procedures in Pulmonary Medicine That Should Be Integrated in a Simulation-Based Curriculum: A

National General Needs Assessment. *Respiration* 2016;91(6):517-22. doi: 10.1159/000446926 [published Online First: 2016/06/12]
483. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Information & Management* 2004;42(1):15-29. doi: https://doi.org/10.1016/j.im.2003.11.002
484. Lindeman B, Brasel K, Minter RM, et al. A Phased Approach: The General Surgery Experience Adopting Entrustable Professional Activities in the United States. *Acad Med* 2021;96(7s):S9-s13. doi: 10.1097/acm.0000000000004107 [published Online First: 2021/06/30]
485. B K, R D, S P, et al. Is the current training pathway for oncoplastic breast surgeons fit for purpose? *The Bulletin of the Royal College of Surgeons of England* 2019;101(2):69-75. doi: 10.1308/rcsbull.2019.69
486. Jensen RD, Paltved C, Jaensch C, et al. Identifying technical skills and clinical procedures in surgery for a simulation-based curriculum: a national general needs assessment. *Surg Endosc* 2022;36(1):47-56. doi: 10.1007/s00464-020-08235-7 [published Online First: 2021/01/06]
487. Szasz P, Louridas M, de Montbrun S, et al. Consensus-based training and assessment model for general surgery. *British Journal of Surgery* 2016;103(6):763-71. doi: 10.1002/bjs.10103
488. Bjerrum F, Strandbygaard J, Rosthøj S, et al. Evaluation of Procedural Simulation as a Training and Assessment Tool in General Surgery-Simulating a Laparoscopic Appendectomy. *J Surg Educ* 2017;74(2):243-50. doi: 10.1016/j.jsurg.2016.08.006 [published Online First: 2016/10/09]
489. Borgersen NJ, Naur TMH, Sørensen SMD, et al. Gathering Validity Evidence for Surgical Simulation: A Systematic Review. *Ann Surg* 2018;267(6):1063-68. doi: 10.1097/sla.0000000000002652 [published Online First: 2018/01/06]
490. Pietersen PI, Bjerrum F, Tolsgaard MG, et al. Standard Setting in Simulation-based Training of Surgical Procedures: A Systematic Review. *Ann Surg* 2022;275(5):872-82. doi: 10.1097/sla.0000000000005209 [published Online First: 2021/09/15]
491. Bjerrum F, Sorensen JL, Konge L, et al. Randomized trial to examine procedure-to-procedure transfer in laparoscopic simulator training. *Br J Surg* 2016;103(1):44-50. doi: 10.1002/bjs.9966 [published Online First: 2015/10/30]
492. Collings AT, Stefanidis D, Doster D, et al. Assessment of Chief Resident Practice Readiness in a Porcine Lab: A 4-Year Experience. *Journal of Surgical Education* 2022;79(3):783-90. doi: https://doi.org/10.1016/j.jsurg.2021.11.009
493. Nayahangan LJ, Clementsen PF, Doubleday A, et al. Developing a simulation-based training curriculum in transesophageal ultrasound with the use of the endobronchial ultrasound-endoscope. *Endosc Ultrasound* 2022;11(2):104-11. doi: 10.4103/eus-d-21-00126 [published Online First: 2022/05/01]
494. Nayahangan LJ, Konge L, Schroeder TV, et al. A National Needs Assessment to Identify Technical Procedures in Vascular Surgery for Simulation Based Training. *Eur J Vasc Endovasc Surg* 2017;53(4):591-99. doi: 10.1016/j.ejvs.2017.01.011 [published Online First: 2017/03/05]
495. Lüscher M, Nayahangan LJ, Thrane JF, et al. Simulation-based surgical training needs in otorhinolaryngology. *Dan Med J* 2022;69(4) [published Online First: 2022/03/24]
496. Thim S, Nayahangan LJ, Paltved C, et al. Identifying and prioritising technical procedures for simulation-based curriculum in paediatrics: a Delphi-based general needs

assessment. *BMJ Paediatr Open* 2020;4(1):e000697. doi: 10.1136/bmjpo-2020-000697 [published Online First: 2020/08/28]
497. Heskin L, Simms C, Traynor O, et al. Designing a synthetic simulator to teach open surgical skills for limb exploration in trauma: a qualitative study exploring the experiences and perspectives of educators and surgical trainees. *BMC Surg* 2021;21(1):417. doi: 10.1186/s12893-021-01417-7 [published Online First: 2021/12/17]
498. Heskin L, Mansour E, Lane B, et al. The impact of a surgical boot camp on early acquisition of technical and nontechnical skills by novice surgical trainees. *Am J Surg* 2015;210(3):570-7. doi: 10.1016/j.amjsurg.2014.12.046 [published Online First: 2015/06/01]
499. Hong M, Rozenblit JW, Hamilton AJ. Simulation-based surgical training systems in laparoscopic surgery: a current review. *Virtual Reality* 2021;25(2):491-510. doi: 10.1007/s10055-020-00469-z
500. Toale C, Morris M, Kavanagh D. Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review. *Am J Surg* 2021;222(4):723-30. doi: 10.1016/j.amjsurg.2021.03.035 [published Online First: 2021/04/15]
501. McDougall EM. Validation of surgical simulators. *J Endourol* 2007;21(3):244-7. doi: 10.1089/end.2007.9985 [published Online First: 2007/04/21]
502. Carter FJ, Schijven MP, Aggarwal R, et al. Consensus guidelines for validation of virtual reality surgical simulators. *Surg Endosc* 2005;19(12):1523-32. doi: 10.1007/s00464-005-0384-2 [published Online First: 2005/10/28]
503. Carter FJ, Schijven MP, Aggarwal R, et al. Consensus guidelines for validation of virtual reality surgical simulators. *Surgical Endoscopy And Other Interventional Techniques* 2005;19(12):1523-32. doi: 10.1007/s00464-005-0384-2
504. Fairhurst K, Strickland A, Maddern G. The LapSim virtual reality simulator: promising but not yet proven. *Surg Endosc* 2011;25(2):343-55. doi: 10.1007/s00464-010-1181-0
505. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine* 2018;169(7):467-73.
506. Smyer P. THE MEDICAL EDUCATION RESEARCH STUDY QUALITY INSTRUMENT (MERSQI)2011.
507. Hambleton RK, Arrasmith G, Sheehan DS, et al. Standards for educational and psychological testing: six reviews: JSTOR, 1986.
508. Doneza JA, Palvia V, Lerner VT, et al. Correlation of surgical case volume and fellowship training with performance on simulated procedural tasks. *Am J Obstet Gynecol* 2021;225(5):548.e1-48.e10. doi: 10.1016/j.ajog.2021.06.062 [published Online First: 2021/06/21]
509. Haidari TA, Bjerrum F, Hansen HJ, et al. Simulation-based VATS resection of the five lung lobes: a technical skills test. *Surg Endosc* 2022;36(2):1234-42. doi: 10.1007/s00464-021-08392-3 [published Online First: 2021/03/05]
510. Jensen K, Hansen HJ, Petersen RH, et al. Evaluating competency in video-assisted thoracoscopic surgery (VATS) lobectomy performance using a novel assessment tool and virtual reality simulation. *Surg Endosc* 2019;33(5):1465-73. doi: 10.1007/s00464-018-6428-1 [published Online First: 2018/09/19]
511. Våpenstad C, Hofstad EF, Bø LE, et al. Lack of transfer of skills after virtual reality simulator training with haptic feedback. *Minim Invasive Ther Allied Technol* 2017;26(6):346-54. doi: 10.1080/13645706.2017.1319866 [published Online First: 2017/05/10]

512. Mathews S, Brodman M, D'Angelo D, et al. Predictors of laparoscopic simulation performance among practicing obstetrician gynecologists. *Am J Obstet Gynecol* 2017;217(5):596.e1-96.e7. doi: 10.1016/j.ajog.2017.07.002 [published Online First: 2017/07/18]
513. Jensen K, Bjerrum F, Hansen HJ, et al. Using virtual reality simulation to assess competence in video-assisted thoracoscopic surgery (VATS) lobectomy. *Surg Endosc* 2017;31(6):2520-28. doi: 10.1007/s00464-016-5254-6 [published Online First: 2016/09/23]
514. Jensen K, Bjerrum F, Hansen HJ, et al. A new possibility in thoracoscopic virtual reality simulation training: development and testing of a novel virtual reality simulator for video-assisted thoracoscopic surgery lobectomy. *Interact Cardiovasc Thorac Surg* 2015;21(4):420-6. doi: 10.1093/icvts/ivv183 [published Online First: 2015/07/15]
515. Alwaal A, Al-Qaoud TM, Haddad RL, et al. Transfer of skills on LapSim virtual reality laparoscopic simulator into the operating room in urology. *Urol Ann* 2015;7(2):172-6. doi: 10.4103/0974-7796.150475 [published Online First: 2015/04/04]
516. Lin D, Pena G, Field J, et al. What are the demographic predictors in laparoscopic simulator performance? *ANZ J Surg* 2016;86(12):983-89. doi: 10.1111/ans.12992 [published Online First: 2015/02/04]
517. Akdemir A, Sendağ F, Oztekin MK. Laparoscopic virtual reality simulator and box trainer in gynecology. *Int J Gynaecol Obstet* 2014;125(2):181-5. doi: 10.1016/j.ijgo.2013.10.018 [published Online First: 2014/03/01]
518. Khan MW, Lin D, Marlow N, et al. Laparoscopic skills maintenance: a randomized trial of virtual reality and box trainer simulators. *J Surg Educ* 2014;71(1):79-84. doi: 10.1016/j.jsurg.2013.05.009 [published Online First: 2014/01/15]
519. Akdemir A, Ergenoğlu AM, Yeniel A, et al. Conventional box model training improves laparoscopic skills during salpingectomy on LapSim: a randomized trial. *J Turk Ger Gynecol Assoc* 2013;14(3):157-62. doi: 10.5152/jtgga.2013.55563 [published Online First: 2013/01/01]
520. Xafis V, Babidge W, Field J, et al. The efficacy of laparoscopic skills training in a Mobile Simulation Unit compared with a fixed site: a comparative study. *Surg Endosc* 2013;27(7):2606-12. doi: 10.1007/s00464-013-2798-6 [published Online First: 2013/02/08]
521. Lehmann KS, Gröne J, Lauscher JC, et al. [Simulation training in surgical education - application of virtual reality laparoscopic simulators in a surgical skills course]. *Zentralbl Chir* 2012;137(2):130-37. doi: 10.1055/s-0031-1283984
522. Tan SC, Marlow N, Field J, et al. A randomized crossover trial examining low- versus high-fidelity simulation in basic laparoscopic skills training. *Surg Endosc* 2012;26(11):3207-14. doi: 10.1007/s00464-012-2326-0 [published Online First: 2012/06/01]
523. Shetty S, Panait L, Baranoski J, et al. Construct and face validity of a virtual reality-based camera navigation curriculum. *J Surg Res* 2012;177(2):191-5. doi: 10.1016/j.jss.2012.05.086 [published Online First: 2012/06/29]
524. Kovac E, Azhar RA, Quirouet A, et al. Construct validity of the LapSim virtual reality laparoscopic simulator within a urology residency program. *Can Urol Assoc J* 2012;6(4):253-59. doi: 10.5489/cuaj.12047
525. Palter VN, Graafland M, Schijven MP, et al. Designing a proficiency-based, content validated virtual reality curriculum for laparoscopic colorectal surgery: A Delphi

approach. *Surgery* 2012;151(3):391-97. doi: https://doi.org/10.1016/j.surg.2011.08.005
526. Våpenstad C, Hofstad E, Bø LE, et al. Limitations of haptic feedback devices on construct validity of the LapSim((R)) virtual reality simulator. *Surg Endosc* 2012;27 doi: 10.1007/s00464-012-2621-9
527. Calatayud D, Arora S, Aggarwal R, et al. Warm-up in a virtual reality environment improves performance in the operating room. *Ann Surg* 2010;251(6):1181-5. doi: 10.1097/SLA.0b013e3181deb630 [published Online First: 2010/05/21]
528. van Dongen KW, Ahlberg G, Bonavina L, et al. European consensus on a competency-based virtual reality training program for basic endoscopic surgical psychomotor skills. *Surg Endosc* 2011;25(1):166-71. doi: 10.1007/s00464-010-1151-6 [published Online First: 2010/06/25]
529. Gauger PG, Hauge LS, Andreatta PB, et al. Laparoscopic simulation training with proficiency targets improves practice and performance of novice surgeons. *Am J Surg* 2010;199(1):72-80. doi: 10.1016/j.amjsurg.2009.07.034 [published Online First: 2010/01/28]
530. Tanoue K, Uemura M, Kenmotsu H, et al. Skills assessment using a virtual reality simulator, LapSim, after training to develop fundamental skills for endoscopic surgery. *Minim Invasive Ther Allied Technol* 2010;19(1):24-9. doi: 10.3109/13645700903492993 [published Online First: 2010/01/26]
531. Maagaard M, Sorensen JL, Oestergaard J, et al. Retention of laparoscopic procedural skills acquired on a virtual-reality surgical trainer. *Surg Endosc* 2011;25(3):722-7. doi: 10.1007/s00464-010-1233-5 [published Online First: 2010/10/12]
532. Schreuder HW, van Dongen KW, Roeleveld SJ, et al. Face and construct validity of virtual reality simulation of laparoscopic gynecologic surgery. *Am J Obstet Gynecol* 2009;200(5):540.e1-8. doi: 10.1016/j.ajog.2008.12.030 [published Online First: 2009/03/17]
533. Larsen CR, Soerensen JL, Grantcharov TP, et al. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. *BMJ* 2009;338:b1802. doi: 10.1136/bmj.b1802
534. Sinha P, Hogle NJ, Fowler DL. Do the laparoscopic skills of trainees deteriorate over time? *Surg Endosc* 2008;22(9):2018-25. doi: 10.1007/s00464-008-9929-5 [published Online First: 2008/04/26]
535. Hogle NJ, Widmann WD, Ude AO, et al. Does training novices to criteria and does rapid acquisition of skills on laparoscopic simulators have predictive validity or are we just playing video games? *J Surg Educ* 2008;65(6):431-5. doi: 10.1016/j.jsurg.2008.05.008 [published Online First: 2008/12/09]
536. Kundhal PS, Grantcharov TP. Psychomotor performance measured in a virtual environment correlates with technical skills in the operating room. *Surg Endosc* 2009;23(3):645-9. doi: 10.1007/s00464-008-0043-5 [published Online First: 2008/07/16]
537. Panait L, Bell RL, Roberts KE, et al. Designing and Validating a Customized Virtual Reality-Based Laparoscopic Skills Curriculum. *Journal of Surgical Education* 2008;65(6):413-17. doi: https://doi.org/10.1016/j.jsurg.2008.08.001
538. Botden SM, Buzink SN, Schijven MP, et al. Augmented versus virtual reality laparoscopic simulation: what is the difference? A comparison of the ProMIS augmented reality laparoscopic simulator versus LapSim virtual reality laparoscopic simulator. *World J*

*Surg* 2007;31(4):764-72. doi: 10.1007/s00268-006-0724-y [published Online First: 2007/03/16]
539. van Dongen KW, Tournoij E, van der Zee DC, et al. Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? *Surg Endosc* 2007;21(8):1413-7. doi: 10.1007/s00464-006-9188-2 [published Online First: 2007/02/13]
540. Newmark J, Dandolu V, Milner R, et al. Correlating virtual reality and box trainer tasks in the assessment of laparoscopic surgical skills. *Am J Obstet Gynecol* 2007;197(5):546.e1-4. doi: 10.1016/j.ajog.2007.07.026 [published Online First: 2007/11/06]
541. Hogle NJ, Briggs WM, Fowler DL. Documenting a learning curve and test-retest reliability of two tasks on a virtual reality training simulator in laparoscopic surgery. *J Surg Educ* 2007;64(6):424-30. doi: 10.1016/j.jsurg.2007.08.007 [published Online First: 2007/12/08]
542. Hassan I, Gerdes B, Koller M, et al. Clinical background is required for optimum performance with a VR laparoscopy simulator. *Comput Aided Surg* 2006;11(2):103-6. doi: 10.3109/10929080600629082 [published Online First: 2006/06/20]
543. Woodrum DT, Andreatta PB, Yellamanchilli RK, et al. Construct validity of the LapSim laparoscopic surgical simulator. *Am J Surg* 2006;191(1):28-32. doi: 10.1016/j.amjsurg.2005.10.018 [published Online First: 2006/01/10]
544. Larsen CR, Grantcharov T, Aggarwal R, et al. Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. *Surg Endosc* 2006;20(9):1460-6. doi: 10.1007/s00464-005-0745-x [published Online First: 2006/07/11]
545. Aggarwal R, Grantcharov TP, Eriksen JR, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. *Annals of surgery* 2006;244(2):310-14. doi: 10.1097/01.sla.0000218094.92650.44
546. Aggarwal R, Tully A, Grantcharov T, et al. Virtual reality simulation training can improve technical skills during laparoscopic salpingectomy for ectopic pregnancy. *Bjog* 2006;113(12):1382-7. doi: 10.1111/j.1471-0528.2006.01148.x [published Online First: 2006/11/04]
547. Eriksen JR, Grantcharov T. Objective assessment of laparoscopic skills using a virtual reality stimulator. *Surg Endosc* 2005;19(9):1216-9. doi: 10.1007/s00464-004-2154-y [published Online First: 2005/09/01]
548. Hassan I, Sitter H, Schlosser K, et al. [A virtual reality simulator for objective assessment of surgeons' laparoscopic skill]. *Chirurg* 2005;76(2):151-6. doi: 10.1007/s00104-004-0936-3 [published Online First: 2004/09/16]
549. Langelotz C, Kilian M, Paul C, et al. LapSim virtual reality laparoscopic simulator reflects clinical experience in German surgeons. *Langenbecks Arch Surg* 2005;390(6):534-7. doi: 10.1007/s00423-005-0571-6 [published Online First: 2005/07/30]
550. Sherman V, Feldman LS, Stanbridge D, et al. Assessing the learning curve for the acquisition of laparoscopic skills on a virtual reality simulator. *Surg Endosc* 2005;19(5):678-82. doi: 10.1007/s00464-004-8943-5 [published Online First: 2005/03/19]
551. Duffy AJ, Hogle NJ, McCarthy H, et al. Construct validity for the LAPSIM laparoscopic surgical simulator. *Surg Endosc* 2005;19(3):401-5. doi: 10.1007/s00464-004-8202-9 [published Online First: 2004/12/30]

552. Lehmann KS, Gröne J, Lauscher JC, et al. Simulation training in surgical education - Application of virtual reality laparoscopic simulators in a surgical skills course. *Zentralblatt fur Chirurgie - Zeitschrift fur Allgemeine, Viszeral- und Gefasschirurgie* 2012;137(2):130-37. doi: 10.1055/s-0031-1283984
553. Hassan I, Sitter H, Schlosser K, et al. Laparoskopiesimulator. *Der Chirurg* 2005;76(2):151-56. doi: 10.1007/s00104-004-0936-3
554. Schreuder HWR, van Dongen KW, Roeleveld SJ, et al. Face and construct validity of virtual reality simulation of laparoscopic gynecologic surgery. *American Journal of Obstetrics and Gynecology* 2009;200(5):540.e1-40.e8. doi: https://doi.org/10.1016/j.ajog.2008.12.030
555. Cosman PH, Hugh TJ, Shearer CJ, et al. Skills acquired on virtual reality laparoscopic simulators transfer into the operating room in a blinded, randomised, controlled trial. *Stud Health Technol Inform* 2007;125:76-81. [published Online First: 2007/03/23]
556. Gauger PG, Hauge LS, Andreatta PB, et al. Laparoscopic simulation training with proficiency targets improves practice and performance of novice surgeons. *The American Journal of Surgery* 2010;199(1):72-80. doi: https://doi.org/10.1016/j.amjsurg.2009.07.034
557. Alwaal A, Al-Qaoud TM, Haddad RL, et al. Transfer of skills on LapSim virtual reality laparoscopic simulator into the operating room in urology. *Urology Annals* 2015;7(2):172-76. doi: 10.4103/0974-7796.150475
558. Zevin B, Aggarwal R, Grantcharov TP. Surgical simulation in 2013: why is it still not the standard in surgical training? *J Am Coll Surg* 2014;218(2):294-301. doi: 10.1016/j.jamcollsurg.2013.09.016 [published Online First: 2013/11/12]
559. Toale C, Morris M, Kavanagh DO. Assessing Operative Skill in the Competency-Based Education Era: Lessons from the UK and Ireland. *Annals of Surgery* 9000 doi: 10.1097/sla.0000000000005242
560. Mayne A, Wilson L, Kennedy N. The Usefulness of Procedure-Based Assessments in Postgraduate Surgical Training Within the Intercollegiate Surgical Curriculum Programme; A Scoping Review. *J Surg Educ* 2020;77(5):1227-35. doi: 10.1016/j.jsurg.2020.03.005 [published Online First: 2020/05/27]
561. Moorthy K, Munz Y, Sarker SK, et al. Objective assessment of technical skills in surgery. *Bmj* 2003;327(7422):1032-7. doi: 10.1136/bmj.327.7422.1032 [published Online First: 2003/11/01]
562. Vaidya A, Aydin A, Ridgley J, et al. Current Status of Technical Skills Assessment Tools in Surgery: A Systematic Review. *J Surg Res* 2020;246:342-78. doi: 10.1016/j.jss.2019.09.006 [published Online First: 2019/11/07]
563. Schijven MP, Reznick RK, ten Cate OTJ, et al. Transatlantic comparison of the competence of surgeons at the start of their professional career. *British Journal of Surgery* 2010;97(3):443-49. doi: 10.1002/bjs.6858
564. Dwyer T, Wright S, Kulasegaram KM, et al. How to set the bar in competency-based medical education: standard setting after an Objective Structured Clinical Examination (OSCE). *BMC Med Educ* 2016;16:1. doi: 10.1186/s12909-015-0506-z [published Online First: 2016/01/06]
565. Andersen SAW, Nayahangan LJ, Park YS, et al. Use of Generalizability Theory for Exploring Reliability of and Sources of Variance in Assessment of Technical Skills: A Systematic Review and Meta-Analysis. *Acad Med* 2021;96(11):1609-19. doi: 10.1097/acm.0000000000004150 [published Online First: 2021/05/06]

566. Kramer A, Muijtjens A, Jansen K, et al. Comparison of a rational and an empirical standard setting procedure for an OSCE. Objective structured clinical examinations. *Med Educ* 2003;37(2):132-9. doi: 10.1046/j.1365-2923.2003.01429.x [published Online First: 2003/02/01]
567. Sousa J, Mansilha A. European panomara on vascular surgery: results from 5 years of FEBVS examinations. *Angiologia e Cirurgia Vascular* 2019;15:171-75. doi: 10.1016/j.ejvs.2021.01.034
568. Gauvin G, Hay K, Hopman W, et al. Competency-based education in general surgery: Are Canadian residents ready? *Canadian Journal of Surgery* 2021;64(5):E473-E75. doi: 10.1503/cjs.011520
569. Diwersi N, Gass J-M, Fischer H, et al. Surgery goes EPA (Entrustable Professional Activity) – how a strikingly easy to use app revolutionizes assessments of clinical skills in surgical training. *BMC Medical Education* 2022;22(1):559. doi: 10.1186/s12909-022-03622-1
570. James HK, Gregory RJH. The dawn of a new competency-based training era. *Bone Jt Open* 2021;2(3):181-90. doi: 10.1302/2633-1462.23.Bjo-2020-0155.R1 [published Online First: 2021/03/20]
571. Borel-Rinkes IHM, Gouma DJ, Hamming JF. Surgical Training in the Netherlands. *World Journal of Surgery* 2008;32(10):2172-77. doi: 10.1007/s00268-007-9460-1
572. Ten Cate O, Khursigara-Slattery N, Cruess RL, et al. Medical competence as a multilayered construct. *Med Educ* 2023 doi: 10.1111/medu.15162 [published Online First: 2023/07/17]
573. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care* 2007;19(6):349-57. doi: 10.1093/intqhc/mzm042
574. Young ME, Ryan A. Postpositivism in Health Professions Education Scholarship. *Acad Med* 2020;95(5):695-99. doi: 10.1097/acm.0000000000003089 [published Online First: 2020/04/30]
575. Burr V, Dick P. Social constructionism. *The Palgrave handbook of critical social psychology* 2017:59-80.
576. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;65(9 Suppl):S63-7. doi: 10.1097/00001888-199009000-00045 [published Online First: 1990/09/01]
577. Andreatta PB, Gruppen LD. Conceptualising and classifying validity evidence for simulation. *Med Educ* 2009;43(11):1028-35. doi: 10.1111/j.1365-2923.2009.03454.x [published Online First: 2009/10/31]
578. Krumm AE, Marcotte K, George BC. Model-Based Operative Performance Expectations for Quantifying Competency in General Surgery. *JAMA Surg* 2023;158(5):515-21. doi: 10.1001/jamasurg.2023.0014 [published Online First: 2023/03/09]
579. Association AER. Validity. *Standards for educational and psychological testing* 2014:11-31.
580. George S, Haque MS, Oyebode F. Standard setting: Comparison of two methods. *BMC Medical Education* 2006;6(1):46. doi: 10.1186/1472-6920-6-46
581. Toale C, Morris M, Kavanagh DO. Training and assessment using the LapSim laparoscopic simulator: a scoping review of validity evidence. *Surg Endosc* 2023;37(3):1658-71. doi: 10.1007/s00464-022-09593-0 [published Online First: 2022/09/20]

582. Vaidya A, Aydin A, Ridgley J, et al. Current Status of Technical Skills Assessment Tools in Surgery: A Systematic Review. *Journal of Surgical Research* 2020;246:342-78. doi: https://doi.org/10.1016/j.jss.2019.09.006
583. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* 2005;15(9):1277-88. doi: 10.1177/1049732305276687 [published Online First: 2005/10/06]
584. Jung JJ, Borkhoff CM, Jüni P, et al. Non-Technical Skills for Surgeons (NOTSS): Critical appraisal of its measurement properties. *Am J Surg* 2018;216(5):990-97. doi: 10.1016/j.amjsurg.2018.02.021 [published Online First: 2018/02/27]
585. Palter VN, Graafland M, Schijven MP, et al. Designing a proficiency-based, content validated virtual reality curriculum for laparoscopic colorectal surgery: a Delphi approach. *Surgery* 2012;151(3):391-7. doi: 10.1016/j.surg.2011.08.005 [published Online First: 2011/10/25]
586. Breimer GE, Haji FA, Bodani V, et al. Simulation-based Education for Endoscopic Third Ventriculostomy: A Comparison Between Virtual and Physical Training Models. *Oper Neurosurg (Hagerstown)* 2017;13(1):89-95. doi: 10.1227/neu.0000000000001317 [published Online First: 2017/09/22]
587. Whitehurst SV, Lockrow EG, Lendvay TS, et al. Comparison of two simulation systems to support robotic-assisted surgical training: a pilot study (Swine model). *J Minim Invasive Gynecol* 2015;22(3):483-8. doi: 10.1016/j.jmig.2014.12.160 [published Online First: 2014/12/30]
588. Alaker M, Wynn GR, Arulampalam T. Virtual reality training in laparoscopic surgery: A systematic review & meta-analysis. *International Journal of Surgery* 2016;29:85-94. doi: https://doi.org/10.1016/j.ijsu.2016.03.034
589. Versteeg M, Hendriks RA, Thomas A, et al. Conceptualising spaced learning in health professions education: A scoping review. *Med Educ* 2020;54(3):205-16. doi: 10.1111/medu.14025 [published Online First: 2019/12/21]
590. Lam K, Chen J, Wang Z, et al. Machine learning for technical skill assessment in surgery: a systematic review. *npj Digital Medicine* 2022;5(1):24. doi: 10.1038/s41746-022-00566-0
591. Soangra R, Sivakumar R, Anirudh ER, et al. Evaluation of surgical skill using machine learning with optimal wearable sensor locations. *PLoS One* 2022;17(6):e0267936. doi: 10.1371/journal.pone.0267936 [published Online First: 2022/06/04]
592. Epstein RM, Hundert EM. Defining and assessing professional competence. *Jama* 2002;287(2):226-35. doi: 10.1001/jama.287.2.226 [published Online First: 2002/01/12]
593. Albanese MA, Mejicano G, Mullan P, et al. Defining characteristics of educational competencies. *Med Educ* 2008;42(3):248-55. doi: 10.1111/j.1365-2923.2007.02996.x [published Online First: 2008/02/16]
594. Parent F, Jouquan J, De Ketele JM. CanMEDS and other "competency and outcome-based approaches" in medical education: clarifying the ongoing ambiguity. *Adv Health Sci Educ Theory Pract* 2013;18(1):115-22. doi: 10.1007/s10459-012-9402-z [published Online First: 2012/09/20]
595. Touchie C, ten Cate O. The promise, perils, problems and progress of competency-based medical education. *Med Educ* 2016;50(1):93-100. doi: 10.1111/medu.12839 [published Online First: 2015/12/24]

596. Glass JM. Competency based training is a framework for incompetence. *Bmj* 2014;348:g2909. doi: 10.1136/bmj.g2909 [published Online First: 2014/04/29]
597. Brooks MA. Medical education and the tyranny of competency. *Perspect Biol Med* 2009;52(1):90-102. doi: 10.1353/pbm.0.0068 [published Online First: 2009/01/27]
598. The surgical curriculum from August 2021 2021 [Available from: https://www.iscp.ac.uk/iscp/curriculum-2021/.
599. James K, Cross K, Lucarotti ME, et al. Undertaking procedure-based assessment is feasible in clinical practice. *Ann R Coll Surg Engl* 2009;91(2):110-2. doi: 10.1308/003588409x359286 [published Online First: 2009/03/26]
600. Brasel KJ, Klingensmith ME, Englander R, et al. Entrustable Professional Activities in General Surgery: Development and Implementation. *Journal of Surgical Education* 2019;76(5):1174-86. doi: https://doi.org/10.1016/j.jsurg.2019.04.003
601. ten Cate O, Khursigara-Slattery N, Cruess RL, et al. Medical competence as a multilayered construct. *Medical Education*;n/a(n/a) doi: https://doi.org/10.1111/medu.15162
602. Carraccio C, Wolfsthal SD, Englander R, et al. Shifting paradigms: from Flexner to competencies. *Acad Med* 2002;77(5):361-7. doi: 10.1097/00001888-200205000-00003 [published Online First: 2002/05/16]
603. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med* 2003;15(4):270-92. doi: 10.1207/s15328015tlm1504_11 [published Online First: 2003/11/13]
604. Govaerts MJ, van der Vleuten CP, Schuwirth LW, et al. Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Adv Health Sci Educ Theory Pract* 2007;12(2):239-60. doi: 10.1007/s10459-006-9043-1 [published Online First: 2006/11/11]
605. Holmboe ES, Sherbino J, Long DM, et al. The role of assessment in competency-based medical education. *Medical Teacher* 2010;32(8):676-82. doi: 10.3109/0142159X.2010.500704
606. Streiner DL, Norman GR, Cairney J, et al. 159Reliability. Health Measurement Scales: A practical guide to their development and use: Oxford University Press 2014:0.
607. Andersen SAW, Park YS, Sørensen MS, et al. Reliable Assessment of Surgical Technical Skills Is Dependent on Context: An Exploration of Different Variables Using Generalizability Theory. *Academic Medicine* 2020;95(12):1929-36. doi: 10.1097/acm.0000000000003550
608. Williams RG, Swanson DB, Fryer JP, et al. How Many Observations are Needed to Assess a Surgical Trainee's State of Operative Competency? *Annals of Surgery* 2019;269(2):377-82. doi: 10.1097/sla.0000000000002554
609. Toale C, Morris M, O'Keeffe D, et al. Assessing operative competence in core surgical training: A reliability analysis. *The American Journal of Surgery* 2023 doi: https://doi.org/10.1016/j.amjsurg.2023.06.020
610. Brydges R, Hatala R, Zendejas B, et al. Linking simulation-based educational assessments and patient-related outcomes: a systematic review and meta-analysis. *Acad Med* 2015;90(2):246-56. doi: 10.1097/acm.0000000000000549 [published Online First: 2014/11/07]
611. Madani A, Vassiliou MC, Watanabe Y, et al. What Are the Principles That Guide Behaviors in the Operating Room?: Creating a Framework to Define and Measure Performance. *Ann Surg* 2017;265(2):255-67. doi: 10.1097/sla.0000000000001962 [published Online First: 2016/09/10]

612. Stefan P, Pfandler M, Kullmann A, et al. Computer-assisted simulated workplace-based assessment in surgery: application of the universal framework of intraoperative performance within a mixed-reality simulation. *BMJ Surg Interv Health Technol* 2023;5(1):e000135. doi: 10.1136/bmjsit-2022-000135 [published Online First: 2023/01/24]
613. Groom P, Schofield L, Hettiarachchi N, et al. Performance of emergency surgical front of neck airway access by head and neck surgeons, general surgeons, or anaesthetists: an in situ simulation study. *British Journal of Anaesthesia* 2019;123(5):696-703. doi: https://doi.org/10.1016/j.bja.2019.07.011
614. Bismilla Z, Boyle T, Mangold K, et al. Development of a Simulation-Based Interprofessional Teamwork Assessment Tool. *J Grad Med Educ* 2019;11(2):168-76. doi: 10.4300/jgme-d-18-00729.1 [published Online First: 2019/04/27]
615. Hack KE, Levy MJ, Garfinkel E, et al. Establishing consensus-based high-acuity low-occurrence skills for EMS physicians: A pilot survey of EMS fellowship faculty. *AEM Education and Training* 2022;6(6):e10828. doi: https://doi.org/10.1002/aet2.10828
616. Bierer J, Horne D, Warren A, et al. Interprofessional Patient-Specific Simulation Preparation to Improve Management of Neonatal High-Acuity Low-Occurrence (HALO) Scenarios. *Canadian Journal of Cardiology* 2021;37(8):1271-74. doi: https://doi.org/10.1016/j.cjca.2021.02.021
617. Hakemi A, Blamoun J, Lundahl A, et al. A Conceptual Framework for Instructional Design of a High Acuity and Low Occurrence Event - Simulation Based Education Training of Residents, Medical Students, and Nurses in Anaphylaxis Utilizing Curated Educational Theories. *Advances in Medical Education and Practice* 2023;14:101-07. doi: 10.2147/AMEP.S398013
618. Fitts PM, Posner MI. Human performance. 1967
619. Boyle E, O'Keeffe DA, Naughton PA, et al. The importance of expert feedback during endovascular simulator training. *J Vasc Surg* 2011;54(1):240-48.e1. doi: 10.1016/j.jvs.2011.01.058 [published Online First: 2011/06/04]
620. Boyle E, Al-Akash M, Gallagher AG, et al. Optimising surgical training: use of feedback to reduce errors during a simulated surgical procedure. *Postgrad Med J* 2011;87(1030):524-8. doi: 10.1136/pgmj.2010.109363 [published Online First: 2011/06/07]
621. Jeannerod M. Mental imagery in the motor context. *Neuropsychologia* 1995;33(11):1419-32. doi: 10.1016/0028-3932(95)00073-c [published Online First: 1995/11/01]
622. Dunnington GL. The art of mentoring. *Am J Surg* 1996;171(6):604-7. doi: 10.1016/s0002-9610(95)00028-3 [published Online First: 1996/06/01]
623. Greenberg JA, Minter RM. Entrustable Professional Activities: The Future of Competency-based Education in Surgery May Already Be Here. *Annals of Surgery* 2019;269(3):407-08. doi: 10.1097/sla.0000000000003153
624. McGaghie WC. Mastery learning: it is time for medical education to join the 21st century. *Acad Med* 2015;90(11):1438-41. doi: 10.1097/acm.0000000000000911 [published Online First: 2015/09/17]
625. McGaghie WC, Siddall VJ, Mazmanian PE, et al. Lessons for continuing medical education from simulation research in undergraduate and graduate medical education: effectiveness of continuing medical education: American College of Chest Physicians

Evidence-Based Educational Guidelines. *Chest* 2009;135(3 Suppl):62s-68s. doi: 10.1378/chest.08-2521 [published Online First: 2009/03/20]
626. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)* 2016;1:31. doi: 10.1186/s41077-016-0033-y [published Online First: 2016/12/07]
627. Amin Z, Boulet JR, Cook DA, et al. Technology-enabled assessment of health professions education: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach* 2011;33(5):364-9. doi: 10.3109/0142159x.2011.565832 [published Online First: 2011/04/27]
628. Andreatta PB, Gruppen LD. Conceptualising and classifying validity evidence for simulation. *Medical Education* 2009;43(11):1028-35. doi: https://doi.org/10.1111/j.1365-2923.2009.03454.x
629. Toale C, Morris M, Ryan DM, et al. Baseline Assessments of Psychomotor Aptitude Are Associated With the Future Operative Performance of Surgical Trainees: A Prospective Observational Cohort Study. *Ann Surg* 2023;278(1):148-52. doi: 10.1097/sla.0000000000005565 [published Online First: 2022/07/16]
630. Buckley C, Kavanagh D, Nugent E, et al. The impact of aptitude on the learning curve for laparoscopic suturing. *American journal of surgery* 2014;207:263-70. doi: 10.1016/j.amjsurg.2013.08.037
631. Buckley C, Kavanagh D, Gallagher T, et al. Does Aptitude Influence the Rate at which Proficiency Is Achieved for Laparoscopic Appendectomy? *Journal of the American College of Surgeons* 2013;217 doi: 10.1016/j.jamcollsurg.2013.07.405
632. Reznick RK, MacRae H. Teaching Surgical Skills — Changes in the Wind. *New England Journal of Medicine* 2006;355(25):2664-69. doi: 10.1056/NEJMra054785
633. Amin Z, Boulet JR, Cook DA, et al. Technology-enabled assessment of health professions education: Consensus statement and recommendations from the Ottawa 2010 conference. *Medical Teacher* 2011;33(5):364-69. doi: 10.3109/0142159X.2011.565832
634. Gardner AK, Dunkin BJ. Pursuing Excellence: The Power of Selection Science to Provide Meaningful Data and Enhance Efficiency in Selecting Surgical Trainees. *Ann Surg* 2019;270(1):188-92. doi: 10.1097/sla.0000000000002806 [published Online First: 2018/05/05]
635. Schreurs S, Cleutjens KB, Muijtjens AMM, et al. Selection into medicine: the predictive validity of an outcome-based procedure. *BMC Medical Education* 2018;18(1):214. doi: 10.1186/s12909-018-1316-x
636. Thomsen ASS, Bach-Holm D, Kjærbo H, et al. Operating Room Performance Improves after Proficiency-Based Virtual Reality Cataract Surgery Training. *Ophthalmology* 2017;124(4):524-31. doi: https://doi.org/10.1016/j.ophtha.2016.11.015
637. Jensen K, Bjerrum F, Hansen HJ, et al. Using virtual reality simulation to assess competence in video-assisted thoracoscopic surgery (VATS) lobectomy. *Surg Endosc* 2017;31(6):2520-28. doi: 10.1007/s00464-016-5254-6
638. Toale C, Morris M, Kavanagh DO. Training to proficiency in surgery using simulation: is there a moral obligation? *J Med Ethics* 2022 doi: 10.1136/medethics-2021-107678 [published Online First: 2022/01/08]
639. Meling TR, Meling TR. The impact of surgical simulation on patient outcomes: a systematic review and meta-analysis. *Neurosurg Rev* 2021;44(2):843-54. doi: 10.1007/s10143-020-01314-2 [published Online First: 2020/05/14]

640. Khan MTA, Patnaik R, Lee CS, et al. Systematic review of academic robotic surgery curricula. *Journal of Robotic Surgery* 2023;17(3):719-43. doi: 10.1007/s11701-022-01500-y
641. Chen R, Rodrigues Armijo P, Krause C, et al. A comprehensive review of robotic surgery curriculum and training for residents, fellows, and postgraduate surgical education. *Surg Endosc* 2020;34(1):361-67. doi: 10.1007/s00464-019-06775-1
642. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int J Surg* 2021;88:105906. doi: 10.1016/j.ijsu.2021.105906 [published Online First: 2021/04/02]
643. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine* 2018;169(7):467-73. doi: 10.7326/M18-0850
644. Artino AR, Jr., La Rochelle JS, Dezee KJ, et al. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach* 2014;36(6):463-74. doi: 10.3109/0142159x.2014.889814 [published Online First: 2014/03/26]
645. Ljuhar D, Nayahangan LJ, Nataraja R, et al. A Bi-national Needs Assessment to Identify and Prioritise Procedures in Paediatric Surgery for Simulation-based Training. *J Pediatr Surg* 2023;58(8):1520-26. doi: 10.1016/j.jpedsurg.2022.10.001 [published Online First: 2022/11/07]
646. Nayahangan LJ, Nielsen KR, Albrecht-Beste E, et al. Determining procedures for simulation-based training in radiology: a nationwide needs assessment. *Eur Radiol* 2018;28(6):2319-27. doi: 10.1007/s00330-017-5244-7 [published Online First: 2018/01/11]
647. Gustavsen PH, Nielsen DG, Paltved C, et al. A national needs assessment study to determine procedures for simulation-based training in cardiology in Denmark. *Scand Cardiovasc J* 2019;53(1):35-41. doi: 10.1080/14017431.2019.1569716 [published Online First: 2019/01/17]
648. Brennan RL. Generalizability Theory. *Educational Measurement: Issues and Practice* 1992;11(4):27-34. doi: https://doi.org/10.1111/j.1745-3992.1992.tb00260.x
649. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach* 2012;34(11):960-92. doi: 10.3109/0142159x.2012.703791 [published Online First: 2012/11/13]

# 8. Supplemental Material

## 8.1 Supplemental Material, Chapter 1

### 8.1.1 Appendix 1.1

Search strategy:

Search terms: (Simulation) AND (Certification OR Selection OR Recruitment OR credentialing OR Progression OR entrustment OR privileging OR autonomy granting) AND (Surgery) AND (internship and residency OR training OR residency education) AND (Assessment)

Search strategy as applied to the Medline (PubMed) database:

Search: **(Simulation) AND (Certification OR Selection OR Recruitment OR credentialing OR Progression OR entrustment OR privileging OR autonomy granting) AND (Surgery) AND (internship and residency OR training OR residency education) AND (Assessment)**
("computer simulation"[MeSH Terms] OR ("computer"[All Fields] AND "simulation"[All Fields]) OR "computer simulation"[All Fields] OR "simulation"[All Fields] OR "simul"[All Fields] OR "simulate"[All Fields] OR "simulated"[All Fields] OR "simulates"[All Fields] OR "simulating"[All Fields] OR "simulation s"[All Fields] OR "simulational"[All Fields] OR "simulations"[All Fields] OR "simulative"[All Fields] OR "simulator"[All Fields] OR "simulator s"[All Fields] OR "simulators"[All Fields]) AND ("certificate"[All Fields] OR "certificate s"[All Fields] OR "certificated"[All Fields] OR "certificates"[All Fields] OR "certification"[MeSH Terms] OR "certification"[All Fields] OR "certifications"[All Fields] OR ("select"[All Fields] OR "selectability"[All Fields] OR "selectable"[All Fields] OR "selected"[All Fields] OR "selecting"[All Fields] OR "selection s"[All Fields] OR "selection, genetic"[MeSH Terms] OR ("selection"[All Fields] AND "genetic"[All Fields]) OR "genetic selection"[All Fields] OR "selection"[All Fields] OR "selectional"[All Fields] OR "selections"[All Fields] OR "selective"[All Fields] OR "selectively"[All Fields] OR "selectives"[All Fields] OR "selectivities"[All Fields] OR "selectivity"[All Fields] OR "selects"[All Fields]) OR ("recruit"[All Fields] OR "recruited"[All Fields] OR "recruiter"[All Fields] OR "recruiters"[All Fields] OR "recruiting"[All Fields] OR "recruitment"[All Fields] OR "recruitments"[All Fields] OR "recruits"[All Fields]) OR ("credentialed"[All Fields] OR "credentialing"[MeSH Terms] OR "credentialing"[All Fields] OR "credential"[All Fields] OR "credentials"[All Fields]) OR ("disease progression"[MeSH Terms] OR ("disease"[All Fields] AND "progression"[All Fields]) OR "disease progression"[All Fields] OR "progression"[All Fields] OR "progress"[All Fields] OR "progressed"[All Fields] OR "progresses"[All Fields] OR "progressing"[All Fields] OR "progressions"[All Fields] OR "progressive"[All Fields] OR "progressively"[All Fields] OR "progressives"[All Fields]) OR ("entrust"[All Fields] OR "entrustability"[All Fields] OR "entrustable"[All Fields] OR "entrusted"[All Fields] OR "entrusting"[All Fields] OR "entrustment"[All Fields] OR "entrusts"[All Fields]) OR ("privilege"[All Fields] OR "privileged"[All Fields] OR

"privileges"[All Fields] OR "privileging"[All Fields]) OR (("autonomies"[All Fields] OR "autonomy"[All Fields]) AND ("financing, organized"[MeSH Terms] OR ("financing"[All Fields] AND "organized"[All Fields]) OR "organized financing"[All Fields] OR "grant"[All Fields] OR "grants"[All Fields] OR "granted"[All Fields] OR "granting"[All Fields]))) AND ("surgery"[MeSH Subheading] OR "surgery"[All Fields] OR "surgical procedures, operative"[MeSH Terms] OR ("surgical"[All Fields] AND "procedures"[All Fields] AND "operative"[All Fields]) OR "operative surgical procedures"[All Fields] OR "general surgery"[MeSH Terms] OR ("general"[All Fields] AND "surgery"[All Fields]) OR "general surgery"[All Fields] OR "surgery s"[All Fields] OR "surgerys"[All Fields] OR "surgeries"[All Fields]) AND ("internship and residency"[MeSH Terms] OR ("internship"[All Fields] AND "residency"[All Fields]) OR "internship and residency"[All Fields] OR ("education"[MeSH Subheading] OR "education"[All Fields] OR "training"[All Fields] OR "education"[MeSH Terms] OR "train"[All Fields] OR "train s"[All Fields] OR "trained"[All Fields] OR "training s"[All Fields] OR "trainings"[All Fields] OR "trains"[All Fields]) OR (("internship and residency"[MeSH Terms] OR ("internship"[All Fields] AND "residency"[All Fields]) OR "internship and residency"[All Fields] OR "residencies"[All Fields] OR "residency"[All Fields] OR "reside"[All Fields] OR "resided"[All Fields] OR "residence"[All Fields] OR "residence s"[All Fields] OR "residences"[All Fields] OR "residency s"[All Fields] OR "resident"[All Fields] OR "resident s"[All Fields] OR "residents"[All Fields] OR "resides"[All Fields] OR "residing"[All Fields]) AND ("educability"[All Fields] OR "educable"[All Fields] OR "educates"[All Fields] OR "education"[MeSH Subheading] OR "education"[All Fields] OR "educational status"[MeSH Terms] OR ("educational"[All Fields] AND "status"[All Fields]) OR "educational status"[All Fields] OR "education"[MeSH Terms] OR "education s"[All Fields] OR "educational"[All Fields] OR "educative"[All Fields] OR "educator"[All Fields] OR "educator s"[All Fields] OR "educators"[All Fields] OR "teaching"[MeSH Terms] OR "teaching"[All Fields] OR "educate"[All Fields] OR "educated"[All Fields] OR "educating"[All Fields] OR "educations"[All Fields]))) AND ("assess"[All Fields] OR "assessed"[All Fields] OR "assessement"[All Fields] OR "assesses"[All Fields] OR "assessing"[All Fields] OR "assessment"[All Fields] OR "assessment s"[All Fields] OR "assessments"[All Fields])

**Translations**

**Simulation:** "computer simulation"[MeSH Terms] OR ("computer"[All Fields] AND "simulation"[All Fields]) OR "computer simulation"[All Fields] OR "simulation"[All Fields] OR "simul"[All Fields] OR "simulate"[All Fields] OR "simulated"[All Fields] OR "simulates"[All Fields] OR "simulating"[All Fields] OR "simulation's"[All Fields] OR "simulational"[All Fields] OR "simulations"[All Fields] OR "simulative"[All Fields] OR "simulator"[All Fields] OR "simulator's"[All Fields] OR "simulators"[All Fields]

**Certification:** "certificate"[All Fields] OR "certificate's"[All Fields] OR "certificated"[All Fields] OR "certificates"[All Fields] OR "certification"[MeSH Terms] OR "certification"[All Fields] OR "certifications"[All Fields]

**Selection:** "select"[All Fields] OR "selectability"[All Fields] OR "selectable"[All Fields] OR "selected"[All Fields] OR "selecting"[All Fields] OR "selection's"[All Fields] OR "selection, genetic"[MeSH Terms] OR ("selection"[All Fields] AND "genetic"[All Fields]) OR "genetic selection"[All Fields] OR "selection"[All Fields] OR "selectional"[All Fields] OR "selections"[All Fields] OR "selective"[All Fields] OR

"selectively"[All Fields] OR "selectives"[All Fields] OR "selectivities"[All Fields] OR "selectivity"[All Fields] OR "selects"[All Fields]
**Recruitment:** "recruit"[All Fields] OR "recruited"[All Fields] OR "recruiter"[All Fields] OR "recruiters"[All Fields] OR "recruiting"[All Fields] OR "recruitment"[All Fields] OR "recruitments"[All Fields] OR "recruits"[All Fields]
**credentialing:** "credentialed"[All Fields] OR "credentialing"[MeSH Terms] OR "credentialing"[All Fields] OR "credential"[All Fields] OR "credentials"[All Fields]
**Progression:** "disease progression"[MeSH Terms] OR ("disease"[All Fields] AND "progression"[All Fields]) OR "disease progression"[All Fields] OR "progression"[All Fields] OR "progress"[All Fields] OR "progressed"[All Fields] OR "progresses"[All Fields] OR "progressing"[All Fields] OR "progressions"[All Fields] OR "progressive"[All Fields] OR "progressively"[All Fields] OR "progressives"[All Fields]
**entrustment:** "entrust"[All Fields] OR "entrustability"[All Fields] OR "entrustable"[All Fields] OR "entrusted"[All Fields] OR "entrusting"[All Fields] OR "entrustment"[All Fields] OR "entrusts"[All Fields]
**privileging:** "privilege"[All Fields] OR "privileged"[All Fields] OR "privileges"[All Fields] OR "privileging"[All Fields]
**autonomy:** "autonomies"[All Fields] OR "autonomy"[All Fields] OR "autonomy's"[All Fields]
**granting:** "financing, organized"[MeSH Terms] OR ("financing"[All Fields] AND "organized"[All Fields]) OR "organized financing"[All Fields] OR "grant"[All Fields] OR "grants"[All Fields] OR "granted"[All Fields] OR "granting"[All Fields]
**Surgery:** "surgery"[Subheading] OR "surgery"[All Fields] OR "surgical procedures, operative"[MeSH Terms] OR ("surgical"[All Fields] AND "procedures"[All Fields] AND "operative"[All Fields]) OR "operative surgical procedures"[All Fields] OR "general surgery"[MeSH Terms] OR ("general"[All Fields] AND "surgery"[All Fields]) OR "general surgery"[All Fields] OR "surgery's"[All Fields] OR "surgerys"[All Fields] OR "surgeries"[All Fields]
**internship and residency:** "internship and residency"[MeSH Terms] OR ("internship"[All Fields] AND "residency"[All Fields]) OR "internship and residency"[All Fields]
**training:** "education"[Subheading] OR "education"[All Fields] OR "training"[All Fields] OR "education"[MeSH Terms] OR "train"[All Fields] OR "train's"[All Fields] OR "trained"[All Fields] OR "training's"[All Fields] OR "trainings"[All Fields] OR "trains"[All Fields]
**residency:** "internship and residency"[MeSH Terms] OR ("internship"[All Fields] AND "residency"[All Fields]) OR "internship and residency"[All Fields] OR "residencies"[All Fields] OR "residency"[All Fields] OR "reside"[All Fields] OR "resided"[All Fields] OR "residence"[All Fields] OR "residence's"[All Fields] OR "residences"[All Fields] OR "residency's"[All Fields] OR "resident"[All Fields] OR "resident's"[All Fields] OR "residents"[All Fields] OR "resides"[All Fields] OR "residing"[All Fields]
**education:** "educability"[All Fields] OR "educable"[All Fields] OR "educates"[All Fields] OR "education"[Subheading] OR "education"[All Fields] OR "educational status"[MeSH Terms] OR ("educational"[All Fields] AND "status"[All Fields]) OR "educational status"[All Fields] OR "education"[MeSH Terms] OR "education's"[All Fields] OR "educational"[All Fields] OR "educative"[All Fields] OR "educator"[All Fields] OR "educator's"[All Fields] OR "educators"[All Fields] OR "teaching"[MeSH

Terms] OR "teaching"[All Fields] OR "educate"[All Fields] OR "educated"[All Fields] OR "educating"[All Fields] OR "educations"[All Fields]
**Assessment:** "assess"[All Fields] OR "assessed"[All Fields] OR "assessement"[All Fields] OR "assesses"[All Fields] OR "assessing"[All Fields] OR "assessment"[All Fields] OR "assessment's"[All Fields] OR "assessments"[All Fields]

## 8.1.2 Appendix 1.2

## Medical Education Research Quality Instrument (MERSQI) scores

| Author | Year | Study design | No of institutions | Response rate | Type of data | Internal structure | Content | Relationships with other variables | Appropriateness of analysis | Complexity of analysis | Outcomes | Total score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lund | 2022 | 1 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 2 | 13 |
| Gangemi | 2022 | 2 | 0.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 14.5 |
| Fong | 2022 | 2 | 1.5 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 9.5 |
| Haidari | 2022 | 2 | 1 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 15 |
| Toale | 2022 | 1 | 1.5 | 1.5 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 9 |
| Hermsen | 2021 | 1.5 | 0.5 | 1.5 | 3 | 0 | 0 | 1 | 1 | 2 | 2 | 12.5 |
| Doneza | 2021 | 1 | 1 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 1.5 | 13 |
| Seaman | 2021 | 2 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 1.5 | 13.5 |
| Engberg | 2021 | 2 | 1.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 15.5 |
| Karani, | 2021 | 1 | 0.5 | 1.5 | 3 | 0 | 0 | 1 | 1 | 2 | 2 | 12 |
| AlJamal | 2020 | 1.5 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 2 | 13.5 |
| Gillis | 2020 | 2 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 1 | 1.5 | 12.5 |
| Kerwin | 2020 | 1 | 1.5 | 1 | 1 | 0 | 1 | 0 | 1 | 2 | 1 | 9.5 |
| Aljamal | 2020 | 2 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 2 | 14 |
| Satava | 2020 | 1 | 1.5 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 7.5 |
| Goderstad | 2020 | 2 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 2 | 14 |
| Halwani | 2019 | 2 | 1.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 15.5 |
| Bube | 2019 | 2 | 1 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 15 |
| Nayahangan | 2019 | 2 | 1.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 2 | 16 |
| Sousa | 2019 | 2 | 1.5 | 1.5 | 3 | 0 | 1 | 1 | 0 | 2 | 1 | 13 |
| Ahmed K | 2018 | 1 | 1.5 | 1.5 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 9 |
| Erturan | 2018 | 2 | 0.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 14.5 |
| Masood | 2018 | 1 | 0.5 | 1.5 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 8 |
| de Montbrun | 2018 | 1 | 1.5 | 1.5 | 1 | 0 | 1 | 0 | 1 | 2 | 1 | 10 |
| Bilgic | 2018 | 2 | 1.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 15.5 |
| de Vries | 2018 | 2 | 1.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 1.5 | 14.5 |
| Moglia | 2018 | 1 | 0.5 | 3 | 1 | 0 | 1 | 1 | 1 | 2 | 2 | 12.5 |
| Hovgaard | 2018 | 2 | 1 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 15 |
| Awad | 2018 | 1 | 1.5 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 6.5 |
| Hazboun | 2017 | 1 | 0.5 | 1.5 | 3 | 0 | 0 | 1 | 1 | 1 | 1.5 | 10.5 |
| Gardner | 2017 | 2 | 1.5 | 1.5 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 10 |
| Louridas | 2017 | 1.5 | 0.5 | 1.5 | 3 | 0 | 1 | 0 | 1 | 2 | 1.5 | 12 |
| Louridas | 2016 | 1 | 1.5 | 1.5 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 8 |
| Winkler-Schwartz | 2016 | 2 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 1.5 | 13.5 |
| Dwyer | 2016 | 2 | 0.5 | 1 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 14 |
| Thinggaard | 2016 | 2 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 1.5 | 13.5 |
| Ritter, | 2016 | 1 | 1.5 | 0 | 1 | 0 | 1 | 0 | 1 | 2 | 1 | 8.5 |
| Janssens | 2015 | 2 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 3 | 15 |
| Thomsen | 2015 | 2 | 0.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 14.5 |
| Deutsch | 2015 | 1 | 1.5 | 0.5 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 7 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jacobsen | 2015 | 2 | 1 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 15 |
| Rooney | 2015 | 2 | 1.5 | 0.5 | 3 | 0 | 1 | 1 | 1 | 2 | 1.5 | 13.5 |
| Groenier | 2015 | 1.5 | 0.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 2 | 14.5 |
| Duran | 2015 | 2 | 0.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 2 | 15 |
| Culligan | 2014 | 2 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 3 | 15 |
| Burden | 2014 | 1 | 1.5 | 1 | 1 | 0 | 1 | 0 | 1 | 2 | 1 | 9.5 |
| Buckley | 2014 | 1.5 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 1.5 | 13 |
| Giannotti | 2014 | 2 | 0.5 | 1.5 | 3 | 0 | 0 | 1 | 1 | 2 | 1.5 | 12.5 |
| Moglia | 2014 | 1 | 1 | 0.5 | 3 | 0 | 1 | 1 | 1 | 2 | 1.5 | 12 |
| Gallagher | 2014 | 1 | 0.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 2 | 14 |
| Ahmed | 2014 | 1 | 1.5 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 7.5 |
| Duran | 2013 | 1 | 1.5 | 1 | 1 | 0 | 0 | 1 | 1 | 2 | 1.5 | 10 |
| Buckley | 2013 | 2 | 0.5 | 1.5 | 3 | 0 | 0 | 1 | 1 | 2 | 1.5 | 12.5 |
| Stefanidis | 2012 | 1 | 1.5 | 0.5 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 8 |
| Rooney | 2012 | 2 | 1.5 | 1.5 | 3 | 0 | 0 | 0 | 1 | 2 | 1.5 | 12.5 |
| Hafford | 2012 | 1.5 | 2.5 | 1.5 | 3 | 0 | 1 | 0 | 1 | 2 | 2 | 14.5 |
| Panait | 2011 | 2 | 0.5 | 1.5 | 3 | 0 | 0 | 1 | 1 | 2 | 1.5 | 12.5 |
| Alvand | 2011 | 3 | 0.5 | 1.5 | 3 | 0 | 1 | 0 | 1 | 2 | 2 | 14 |
| Van Herzeele | 2010 | 1 | 1 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 2 | 13.5 |
| Xeroulis | 2009 | 2 | 1.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 15.5 |
| Carroll | 2009 | 1 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 1.5 | 12.5 |
| Salgado | 2009 | 1 | 0.5 | 1.5 | 3 | 1 | 0 | 1 | 1 | 2 | 1.5 | 12.5 |
| Gallagher | 2008 | 1 | 0.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 13.5 |
| Tedesco | 2008 | 1 | 0.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 1.5 | 12.5 |
| Goldberg | 2008 | 1 | 0.5 | 1 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 13 |
| Cope | 2008 | 1 | 1 | 1.5 | 3 | 0 | 0 | 1 | 1 | 2 | 1.5 | 12 |
| Tansley | 2007 | 1 | 1.5 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 7.5 |
| Aggarwal | 2006 | 2 | 1.5 | 1.5 | 3 | 0 | 1 | 1 | 1 | 1 | 1.5 | 13.5 |
| Pandey | 2006 | 1 | 1.5 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 2 | 15 |

*Medical Education Research Study Quality Index (MERSQI) score for included articles in Chapter 1: The current status of high-stakes technical performance assessments in surgical training. MERSQI items available at: Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. JAMA. 2007 Sep 5;298(9):1002-9. doi: 10.1001/jama.298.9.1002. PMID: 17785645.*

## 8.2 Supplemental Material – Chapter 2

**8.2.1 Appendix 2.2.1** Search strategy as applied to the MEDLINE (PubMed) database – Assessing operative skill in the competency-based education era, lessons from the UK and Ireland.

Available at: Toale C, Morris M, Kavanagh DO. Assessing Operative Skill in the Competency-based Education Era: Lessons From the UK and Ireland. Ann Surg. 2022 Apr 1;275(4):e615-e625. doi: 10.1097/SLA.0000000000005242. PMID: 35129531.

### 8.2.2 Appendix 2.2.2:

Quality assessment of included studies according to the National Institute of Health/National Heart, Lung and Blood Institute (NIH/NHLBI) Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies (available at https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools). Assessment of Operative Skill in the Competency-Based Education Era – Lessons from the UK and Ireland

Available at: Toale C, O'Byrne A, Morris M, Kavanagh DO. Defining operative experience targets in surgical training: A systematic review. Surgery. 2022 Nov;172(5):1364-1372. doi: 10.1016/j.surg.2022.07.015. Epub 2022 Aug 26. PMID: 36038374.

### 8.2.3 Appendix 2.3.1:

Search strategy as applied to the Medline (Pubmed) database – Defining Operative Experience Targets in Surgical Training: A Systematic Review

Available at: Toale C, O'Byrne A, Morris M, Kavanagh DO. Defining operative experience targets in surgical training: A systematic review. Surgery. 2022 Nov;172(5):1364-1372. doi: 10.1016/j.surg.2022.07.015. Epub 2022 Aug 26. PMID: 36038374.

### 8.2.4 Appendix 2.3.2:

Data Extraction Pro-Forma - Defining Operative Experience Targets in Surgical Training: A Systematic Review

Available at: Toale C, O'Byrne A, Morris M, Kavanagh DO. Defining operative experience targets in surgical training: A systematic review. Surgery. 2022 Nov;172(5):1364-1372. doi: 10.1016/j.surg.2022.07.015. Epub 2022 Aug 26. PMID: 36038374.

### 8.2.5 Appendix 2.3.3.

Available at: Toale C, O'Byrne A, Morris M, Kavanagh DO. Defining operative experience targets in surgical training: A systematic review. Surgery. 2022 Nov;172(5):1364-1372. doi: 10.1016/j.surg.2022.07.015. Epub 2022 Aug 26. PMID: 36038374.

### 8.2.6 Appendix 2.4.1:

Available at: Toale C, Morris M, Kavanagh D. Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review. Am J Surg. 2021 Oct;222(4):723-730. doi: 10.1016/j.amjsurg.2021.03.035. Epub 2021 Apr 6. PMID: 33849711.

### 8.2.7 Appendix 2.4.2

Available at: Toale C, Morris M, Kavanagh D. Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review. Am J Surg. 2021 Oct;222(4):723-730. doi: 10.1016/j.amjsurg.2021.03.035. Epub 2021 Apr 6. PMID: 33849711.

**8.2.8 Appendix 2.5.1.** Semi-Structured Interview Guide **-** Perspectives on Simulation-Based Assessment of Operative Skill in Surgical Training

Available at: Toale C, Morris M, Kavanagh DO. Perspectives on simulation-based assessment of operative skill in surgical training. Med Teach. 2022 Oct 26:1-8. doi: 10.1080/0142159X.2022.2134001. Epub ahead of print. PMID: 36288727.

**8.2.9 Appendix 2.5.2:** Themes, Sub-Themes, Codes, and Example Quotes – Perspectives on Simulation-Based Assessment of Operative Skill in Surgical Training

Available at: Toale C, Morris M, Kavanagh DO. Perspectives on simulation-based assessment of operative skill in surgical training. Med Teach. 2022 Oct 26:1-8. doi: 10.1080/0142159X.2022.2134001. Epub ahead of print. PMID: 36288727.

### 8.2.10 Appendix 2.6.1

Survey Items - The Acceptability of Assessing Operative Competence Outside of the Operating Room: A Survey of Surgical Trainees

**Appendix 1: Survey Items**

---

**1. What is your current level of training?**

| | |
|---|---|
| ST 1 | |
| ST 2 | |
| ST 3 | |
| ST 4 | |
| ST 5 | |
| ST 6 | |
| ST 7 | |
| ST 8 | |
| Out of programme research | |
| Post CCT- Fellowship | |
| Consultant trainer | |
| Other | |

**2. What sub-specialty is your current training post/consultancy post within?**

| | |
|---|---|
| Cardiothoracic Surgery | |
| General Surgery | |
| Plastic and Reconstructive Surgery | |
| Trauma and Orthopaedics | |
| Urology | |
| Vascular Surgery | |
| Otolaryngology/ Head and Neck Surgery | |
| Neurosurgery | |
| Oral and Maxillo-Facial Surgery | |
| Paediatric Surgery | |

**3. Which of the following types of simulator model have you used for learning, teaching or assessment? (select all that apply)**

| | |
|---|---|
| Low-fidelity synthetic bench models (e.g. laparoscopic box trainers or bench-model suturing kits) | |
| High-fidelity synthetic bench models (synthetic replicas of anatomical models for surgical procedures) | |
| Animal models | |
| Virtual reality or computer- generated models (e.g. computer generated laparoscopic/ endovascular/ robotic models) | |
| Cadaveric models | |

**Please rate your responses to the following statements from 1 – 6, where 1 = strongly disagree, 2 = disagree, 3 = somewhat disagree, 4 = somewhat agree, 5 = agree, 6 = strongly agree**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 4 | I am confident that when I obtain my certificate of completion of training (CCT) I will be competent in performing the expected operative procedures in my specialty without assistance, including complications | 1 | 2 | 3 | 4 | 5 | 6 |
| 5 | I am confident that I will have had sufficient operative experience upon obtaining my CCT to ensure that I will be competent in performing the expected operative procedures in my specialty without assistance, including when managing complications | 1 | 2 | 3 | 4 | 5 | 6 |
| 6 | Current methods of operative competency assessment (procedure-based assessments, multiple consultant reports, multi-source feedback and indicative number targets) are satisfactory as summative assessments of operative competence | 1 | 2 | 3 | 4 | 5 | 6 |
| 7 | Using simulation to assess operative competence is more objective than assessments performed in the operating theatre | 1 | 2 | 3 | 4 | 5 | 6 |
| 8 | Using simulation to assess operative competence is easier to standardise than assessments performed in the operating theatre | 1 | 2 | 3 | 4 | 5 | 6 |
| 9 | Simulation-based assessment is a more constructive way of receiving feedback or giving feedback to trainees than in-theatre assessments | 1 | 2 | 3 | 4 | 5 | 6 |
| 10 | Using simulation-based assessment to ensure trainees meet the minimum level of operative competency expected for their stage of training would allow in-theatre training time to be used more efficiently | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 11 | A trainee's performance in a simulated operative procedure is reflective of their performance in the operating theatre | 1 | 2 | 3 | 4 | 5 | 6 |
| 12 | Simulation is an accurate method of assessing whether a trainee knows the steps of a procedure | 1 | 2 | 3 | 4 | 5 | 6 |
| 13 | Simulation is an accurate method of assessing a trainee's operative technique or skill | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 14 | Simulation is an accurate method of assessing a trainee's non-technical skills such as communication and teamwork | 1 | 2 | 3 | 4 | 5 | 6 |
| 15 | Simulation is an accurate method of assessing a trainee's ability to perform an operative procedure under pressure | 1 | 2 | 3 | 4 | 5 | 6 |
| 16 | Simulation is an accurate method of assessing a trainee's intra-operative decision-making skills | 1 | 2 | 3 | 4 | 5 | 6 |
| 17 | Simulation-based assessment could be used to determine whether trainees meet the minimum operative competency level expected for their stage of training | 1 | 2 | 3 | 4 | 5 | 6 |
| 18 | Simulation-based assessment of operative competence can identify under-performing trainees | 1 | 2 | 3 | 4 | 5 | 6 |
| 19 | Simulation-based assessments can distinguish between trainees who meet operative competency expectations and trainees who exceed expectations | 1 | 2 | 3 | 4 | 5 | 6 |
| 20 | Using simulation-based assessment to ensure trainees meet the minimum operative competency level expected for their stage of training would have a positive impact on patient safety | 1 | 2 | 3 | 4 | 5 | 6 |
| 21 | Simulation-based assessment can accurately replicate the real-world operating theatre environment | 1 | 2 | 3 | 4 | 5 | 6 |
| 22 | Performance in simulation-based assessment can provide objective evidence of a trainee's progression in operative competence over time | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 23 | A human assessor is a better determinant of a trainee's ability than simulator-measured metrics (i.e. metrics recorded by a computer) | 1 | 2 | 3 | 4 | 5 | 6 |
| 24 | Simulator-measured metrics , recorded and assessed by a computer, are accurate methods of assessing operative competence | 1 | 2 | 3 | 4 | 5 | 6 |
| 25 | 'Logged hours' spent using a simulator is reflective of a trainee's operative competence | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 26 | If simulation were to be used to assess a trainee's operative competence, **the models would need to have a very realistic appearance** | 1 | 2 | 3 | 4 | 5 | 6 |
| 27 | If simulation were to be used to assess a trainee's operative competence, **the models would have to reflect real-world case variety** | 1 | 2 | 3 | 4 | 5 | 6 |
| 28 | If simulation were to be used to assess a trainee's operative competence, **it would be important for the simulation to reflect the real-world theatre environment** | 1 | 2 | 3 | 4 | 5 | 6 |
| 29 | If simulation were to be used to assess a trainee's operative competence, **it would be important for the simulated models to handle like human tissue** | 1 | 2 | 3 | 4 | 5 | 6 |
| 30 | The more senior the trainee being assessed, the more realistic the simulator model should be | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 31 | **Low-fidelity synthetic bench models** (e.g. laparoscopic box trainers or bench-model suturing kits) are suitable for use in summative assessments of a trainee's operative competence | 1 | 2 | 3 | 4 | 5 | 6 |
| 32 | **High-fidelity synthetic bench models** (synthetic replicas of anatomical models for surgical procedures) are suitable for use in summative assessments of a trainee's operative competence | 1 | 2 | 3 | 4 | 5 | 6 |
| 33 | **Animal models** are suitable for use in summative assessments of a trainee's operative competence | 1 | 2 | 3 | 4 | 5 | 6 |
| 34 | **Virtual reality or computer-generated platforms** (e.g. robotic/ laparoscopic/ endovascular platforms) are suitable for use in summative assessments of a trainee's operative competence | 1 | 2 | 3 | 4 | 5 | 6 |
| 35 | **Cadaveric models** are suitable for use in summative assessments of a trainee's operative competence | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 36 | Simulation-based assessment is suitable for assessing a trainee's competence in **open surgery** | 1 | 2 | 3 | 4 | 5 | 6 |
| 37 | Simulation-based assessment is suitable for assessing a trainee's competence in **Laparoscopic surgery** | 1 | 2 | 3 | 4 | 5 | 6 |
| 38 | Simulation-based assessment is suitable for assessing a trainee's competence in **Endoscopic surgery** | 1 | 2 | 3 | 4 | 5 | 6 |
| 39 | Simulation-based assessment is suitable for assessing a trainee's competence in **Endovascular surgery** | 1 | 2 | 3 | 4 | 5 | 6 |
| 40 | Simulation-based assessment is suitable for assessing a trainee's competence in **Robotic surgery** | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 41 | If simulation were to be used to assess operative competence, it would suffice to simulate key steps of the procedure, and not necessarily the whole procedure | 1 | 2 | 3 | 4 | 5 | 6 |
| 42 | Simulation should be used to assess competence on a series of core, commonly encountered procedures | 1 | 2 | 3 | 4 | 5 | 6 |
| 43 | Simulation should be used to assess competence in rarely encountered, high-risk procedures | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 44 | Technical ability is an important consideration when selecting trainees for higher specialist training | 1 | 2 | 3 | 4 | 5 | 6 |
| 45 | Using simulation-based operative assessment to inform selection decisions for higher specialist training is fair | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 46 | A trainee's performance in simulation-based assessment should be used to decide how much autonomy trainees are granted in the operating theatre | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 47 | Trainees should have to demonstrate competence on a simulated procedure before they can perform that procedure on a patient under supervision as the primary operator | 1 | 2 | 3 | 4 | 5 | 6 |
| 48 | Using simulation-based assessment to decide how much autonomy trainees are granted in the operating room would result in less operative exposure for trainees | 1 | 2 | 3 | 4 | 5 | 6 |
| 49 | Using simulation-based assessment to decide how much autonomy trainees are granted in the operating theatre would positively incentivize trainees to improve their operative performance | 1 | 2 | 3 | 4 | 5 | 6 |
| 50 | Using simulation-based assessment to decide how much autonomy trainees are granted in the operating room would positively impact upon patient safety | 1 | 2 | 3 | 4 | 5 | 6 |
| 51 | Using simulation-based assessment to decide how much autonomy trainees are granted in the operating room would improve trainee confidence | 1 | 2 | 3 | 4 | 5 | 6 |
| 52 | Simulation-based assessment should replace a trainer's expert opinion regarding the appropriate amount of autonomy to give to a trainee | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 53 | It is fair to use scores in simulation-based assessment, in conjunction with other assessments, to inform decisions on a trainee's progression to the next training year | 1 | 2 | 3 | 4 | 5 | 6 |
| 54 | Performance in simulation-based assessment alone is insufficient when making decisions regarding a trainee's progression to the next training year | 1 | 2 | 3 | 4 | 5 | 6 |
| 55 | It is more difficult to assess the operative competence of more senior trainees using simulation | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 56 | A trainee's operative competence should be formally assessed at the end of training | 1 | 2 | 3 | 4 | 5 | 6 |
| 57 | Simulation-based assessment could be used to inform decisions regarding a trainee's operative competence at the end of training, in conjunction with other assessments | 1 | 2 | 3 | 4 | 5 | 6 |
| 58 | It is more ethical, from a patient's perspective, to use simulation-based assessment than in-theatre operative cases to assess trainees at the end of training | 1 | 2 | 3 | 4 | 5 | 6 |
| 59 | Using simulation-based assessment to assess a trainee's operative competence at the end of training is fair | 1 | 2 | 3 | 4 | 5 | 6 |
| 60 | Using simulation-based assessment alone is sufficient when making decisions regarding a trainee's operative competence at the end of training | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 61 | Simulators **that I have experienced** are suitable for use in simulation-based assessments informing decisions regarding selecting trainees for higher specialist training | 1 | 2 | 3 | 4 | 5 | 6 |
| 62 | Simulators **that I have experienced**, are suitable for use in informing decisions regarding how much autonomy trainees are granted in the operating theatre | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 63 | Simulators **that I have experienced** are suitable for use in informing decisions regarding the progression of trainees through higher specialist training | 1 | 2 | 3 | 4 | 5 | 6 |
| 64 | Simulators **that I have experienced** are suitable for use in assessing a trainee's operative competence at the end of training | 1 | 2 | 3 | 4 | 5 | 6 |

65. What should the consequences of failing a simulation-based assessment of operative competence be? (select all that apply)

| | |
|---|---|
| No consequence | |
| Reduced operative autonomy | |
| Dedicated in theatre operative teaching followed by re-assessment | |
| Simulation-based teaching followed by re-assessment | |
| Further re-assessment without mandated teaching time | |
| Structured feedback without further assessment | |
| Failure to progress to the next training year or failure to achieve CCT | |
| Other (please elaborate) | |

66. What concerns do you have regarding the use of simulation-based assessment of operative competence in surgery?

| |
|---|
| |

67. What are the positive aspects of simulation-based assessment of operative competence in surgery?

| |
|---|
| |

## 8.3 Supplemental Material – Chapter 3

### 8.3.1 Appendix 3.1.1

RCSI SURGICAL AFFAIRS

**NATIONAL SURGICAL TRAINING CENTRE**

**Supervised Structured Assessment of Operative Performance (SSAOP)**

RCSI DEVELOPING HEALTHCARE LEADERS WHO MAKE A DIFFERENCE WORLDWIDE

| | | |
|---|---|---|
| **Trainee Name:** | **Trainer Name:** | |
| **Hospital:** | **Specialty:** | |
| **Programme Year:** | **Rotation Start Date:** | **End Date:** |

**Procedure Performed:**

This is an assessment of operative performance of the surgical trainee. Trainees must complete three SSAOPs in each six month rotation (i.e. 12 in total during CST). Each SSAOP should be for a different surgical procedure. The assessment can be performed on any surgical procedure which the trainer considers appropriate for the specialty and the trainee year of training e.g. appendectomy, or repair of inguinal hernia. Minor operations (e.g. excision of skin lesions) are unacceptable.

**NB. The completed form will NOT be considered if "N/A" is entered for three or more of the items numbered 1 -15 (i.e. 20% of the items)**

**A maximum of 15 points is awarded per SSAOP. Each SSAOP is scored out of 150 for ease of calculation.**

**N/A = not observed or not done by the trainee**

**Please mark all sections by checking the box in the appropriate category.**

**Case Difficulty (Max 15 marks)**

| Straightforward anatomy, no related prior operation(s) or treatment (e.g. radiation) (5) | Intermediate difficulty (10) | Obesity, abnormal anatomy, complex pathology, related prior operation(s) or treatment (15) |
|---|---|---|
| ☐ | ☐ | ☐ |

**Degree of Prompting or Direction (Max 15 marks)**

| Substantial direction by consultant. Trainee performs all steps but the consultant provides constant direction to the trainee and surgical team. (5) | Some direction by consultant. Trainee performs all steps but the consultant provides occasional direction to the trainee and /or the surgical team. (10) | Minimal direction by consultant. Trainee performs all steps and directs the surgical team independently with little or no direction from the consultant to either the trainee or the surgical team. (15) |
|---|---|---|
| ☐ | ☐ | ☐ |

| **Pre-Operative Preparation (Max 30 marks)**<br>***Please indicate the degree to which trainee:*** | Very poor. Unacceptable for level of training<br>(1) | Below expectations for level of training<br>(2) | Meets expectations for level of training<br>(3) | Above expectations for level of training<br>(4) | Exceptional. Capable of performing independently<br>(5) | N/A |
|---|---|---|---|---|---|---|
| 1. Demonstrates knowledge of case & relevant investigations. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Ensures availability of relevant patient data & x-rays. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Ensures patient is appropriately positioned with attention to pressure areas & diathermy pad. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Initiates "Time out" in accordance with WHO guidelines | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5.Demonstrates effective scrubbing & gowning technique | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Demonstrates appropriate skin preparation & draping. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| **Operative Procedure (Max 80 marks).**<br>***Please indicate the degree to which trainee:*** | Very poor. Unacceptable for level of training<br>(2) | Below expectations for level of training<br>(4) | Meets expectations for level of training<br>(6) | Above expectations for level of training<br>(8) | Exceptional. Capable of performing independently<br>(10) | N/A |
|---|---|---|---|---|---|---|
| 7. Plans & makes incision(s) taking into account procedure & prior surgery. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. Chooses appropriate surgical instruments (including sutures). | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. Uses surgical instruments (includes diathermy, stapler etc.) in a manner that is safe (i.e. visualizes tips) & demonstrates economy of motion. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. Handles tissue carefully, demonstrates meticulous dissection technique & appropriate use of tissue planes. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 11. Anticipates needs, thinks ahead & sets up operative field by effective use of retraction & assistants, with efficient forward progression & flow. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 12. Demonstrates a calm demeanor, is not distracted, slows down when appropriate & accepts feedback. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | | | | | | |
|---|---|---|---|---|---|---|
| **13. Is able to problem solve, recognise, avoid & recover from errors & unexpected events (includes abnormal anatomy & bleeding).** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| **14. Communicates effectively with team, provides leadership & appropriately monitors changes outside immediate operative field.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| **Postoperative Care (10)** *Please indicate the degree to which the trainee:* | Very poor. Unacceptable for level of training (2) | Below expectations for level of training (4) | Meets expectations for level of training (6) | Above expectations for level of training (8) | Exceptional. Capable of performing independently (10) | N/A |
|---|---|---|---|---|---|---|
| **15. Completes the operation note & writes post-operative orders.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

**Please indicate the weaknesses in this trainee's performance:**

**Please indicate the strengths in this trainee's performance:**

Please rate this trainee's overall performance in this procedure. Exceptional rating indicates technically proficient performance (i.e., trainee is ready to perform **THIS** operation independently assuming he/she consistently performs at this level).

| **Unable to perform basic operative steps without direction** | **Fair performance but needed considerable direction to complete steps.** | **Good performance but needed some direction** | **Excellent performance but needed some direction** | **Exceptional performance. Could operate independently.** |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

Assessor's Signature: _____

Assessor's Name _____ (Print)

Date _____

Office Use Only

Score: _____

### 8.3.2 Appendix 3.2.1 : Assessment Tools

RCSI SURGICAL AFFAIRS

# NATIONAL SURGICAL TRAINING CENTRE

**Supervised Structured Assessment of Operative Performance (SSAOP)**

RCSI DEVELOPING HEALTHCARE LEADERS WHO MAKE A DIFFERENCE WORLDWIDE

| | | |
|---|---|---|
| **Trainee Name:** | **Trainer Name:** | |
| **Hospital:** | **Specialty:** | |
| **Programme Year:** | **Rotation Start Date:** | **End Date:** |

**Procedure Performed:**

This is an assessment of operative performance of the surgical trainee. Trainees must complete three SSAOPs in each six month rotation (i.e. 12 in total during CST). Each SSAOP should be for a different surgical procedure. The assessment can be performed on any surgical procedure which the trainer considers appropriate for the specialty and the trainee year of training e.g. appendectomy, or repair of inguinal hernia. Minor operations (e.g. excision of skin lesions) are unacceptable.

**NB. The completed form will NOT be considered if "N/A" is entered for three or more of the items numbered 1 -15 (i.e. 20% of the items)**

**A maximum of 15 points is awarded per SSAOP. Each SSAOP is scored out of 150 for ease of calculation.**

**N/A = not observed or not done by the trainee**

**Please mark all sections by checking the box in the appropriate category.**

**Case Difficulty (Max 15 marks)**

| | | |
|---|---|---|
| Straightforward anatomy, no related prior operation(s) or treatment (e.g. radiation)<br>(5) | Intermediate difficulty<br>(10) | Obesity, abnormal anatomy, complex pathology, related prior operation(s) or treatment<br>(15) |
| ☐ | ☐ | ☐ |

**Degree of Prompting or Direction (Max 15 marks)**

| | | |
|---|---|---|
| Substantial direction by consultant. Trainee performs all steps but the consultant provides constant direction to the trainee and surgical team.<br>(5) | Some direction by consultant. Trainee performs all steps but the consultant provides occasional direction to the trainee and /or the surgical team.<br>(10) | Minimal direction by consultant. Trainee performs all steps and directs the surgical team independently with little or no direction from the consultant to either the trainee or the surgical team.<br>(15) |
| ☐ | ☐ | ☐ |

| **Pre-Operative Preparation (Max 30 marks)**<br>***Please indicate the degree to which trainee:*** | Very poor. Unacceptable for level of training<br>(1) | Below expectations for level of training<br>(2) | Meets expectations for level of training<br>(3) | Above expectations for level of training<br>(4) | Exceptional. Capable of performing independently<br>(5) | N/A |
|---|---|---|---|---|---|---|
| 1. Demonstrates knowledge of case & relevant investigations. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Ensures availability of relevant patient data & x-rays. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Ensures patient is appropriately positioned with attention to pressure areas & diathermy pad. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Initiates "Time out" in accordance with WHO guidelines | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5.Demonstrates effective scrubbing & gowning technique | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Demonstrates appropriate skin preparation & draping. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| **Operative Procedure (Max 80 marks).**<br>***Please indicate the degree to which trainee:*** | Very poor. Unacceptable for level of training<br>(2) | Below expectations for level of training<br>(4) | Meets expectations for level of training<br>(6) | Above expectations for level of training<br>(8) | Exceptional. Capable of performing independently<br>(10) | N/A |
|---|---|---|---|---|---|---|
| 7. Plans & makes incision(s) taking into account procedure & prior surgery. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. Chooses appropriate surgical instruments (including sutures). | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. Uses surgical instruments (includes diathermy, stapler etc.) in a manner that is safe (i.e. visualizes tips) & demonstrates economy of motion. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. Handles tissue carefully, demonstrates meticulous dissection technique & appropriate use of tissue planes. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 11. Anticipates needs, thinks ahead & sets up operative field by effective use of retraction & assistants, with efficient forward progression & flow. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 12. Demonstrates a calm demeanor, is not distracted, slows down when appropriate & accepts feedback. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | | | | | | |
|---|---|---|---|---|---|---|
| **13. Is able to problem solve, recognise, avoid & recover from errors & unexpected events (includes abnormal anatomy & bleeding).** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| **14. Communicates effectively with team, provides leadership & appropriately monitors changes outside immediate operative field.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| **Postoperative Care (10)** *Please indicate the degree to which the trainee:* | Very poor. Unacceptable for level of training (2) | Below expectations for level of training (4) | Meets expectations for level of training (6) | Above expectations for level of training (8) | Exceptional. Capable of performing independently (10) | N/A |
|---|---|---|---|---|---|---|
| **15. Completes the operation note & writes post-operative orders.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

**Please indicate the weaknesses in this trainee's performance:**

**Please indicate the strengths in this trainee's performance:**

Please rate this trainee's overall performance in this procedure. Exceptional rating indicates technically proficient performance (i.e., trainee is ready to perform **THIS** operation independently assuming he/she consistently performs at this level).

| **Unable to perform basic operative steps without direction** | **Fair performance but needed considerable direction to complete steps.** | **Good performance but needed some direction** | **Excellent performance but needed some direction** | **Exceptional performance. Could operate independently.** |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

Assessor's Signature: _____

Assessor's Name _____ (Print)

Date _____

Office Use Only

Score: _____

RCSI

# Operative Skills Score Sheet 2018

**Assessor Name (legible):** _____

**IMC number for CPD:** _____ **Candidate ID:**

**Date:** _____ **Time:** _____

Procedure: **BOWEL ANASTOMOSIS**

Instructions: **Perform an end to end bowel anastomosis using a single layer of interrupted inverting sutures, hand tied. You may request assistance if required.**

| | Not Done | Done |
|---|---|---|
| **Alignment of bowel ends** | | |
| 1.Candidate aligns bowel ends correctly (mesenteric to mesenteric side) | | |
| 2. Secures corners (three o'clock and nine o'clock) with stay sutures | | |
| | | |
| **Suture placement** | | |
| 4.Places adequate number of sutures (i.e., every 3mm approx: > every 5mm = incorrect) | | |
| 5. Sutures appropriate distance from bowel edge (ie. ~3 mm; >4mm = incorrect) | | |
| 6. Sutures placed at equal intervals and parallel to each other | | |
| **Knot tying** | | |
| 7. Hand-tied knots with minimum 3 throws | | |
| 8. All throws square | | |
| **Apposition of bowel** | | |
| 9. Sutures tied at correct tension | | |
| 10. Bowel evenly apposed (without dog ears i.e., puckered area of bowel between two stitches) | | |
| 11. Bowel edges inverted | | |
| 12. All knots on the inside (except final knot which may be external) | | |
| 12. Gap-free anastomosis ( likely to be water-tight in vivo) | | |
| 13. Gentle tissue handling throughout | | |
| **Total** | | |

| **What is the overall competence of this candidate? (independent of the scores above)** | | |
|---|---|---|
| **Not Competent/Fail** ☐ | **Borderline** ☐ | **Competent/Pass** ☐ |

**Comments:**

**PLEASE TURN OVER TO COMPLETE DETAILED GLOBAL RATING PAGE** ➡

*Updated April 2018*

RCSI

## TRAINEE ASSESSMENT REPORT (TAR)

| | |
|---|---|
| **Trainee Name:** | **Trainer Name:** |
| **Hospital:** | **Specialty:** |
| **Year:** | **Rotation Date:** |

**Please note the TAR is completed by your current Trainer or those trainers you worked with in the recent past as listed on your clinical experience.**

**Performance Ratings Scale**

1. **Very Poor** – Performs below standards and expectations / warrants immediate performance management / remediation action.
2. **Below Expectations** – Has met some performance objectives, but overall failed to meet performance standards on some key dimensions.
3. **Meets Expectations** – Successfully meets performance standards and expectations.
4. **Above Expectations** – Performance consistently meets and sometimes exceeds standards and expectations. Performance is frequently above expectation.
5. **Exceptional** – Performance consistently exceeds all standards and expectations, in conjunction with all areas of personal skills, relationship management and communications skills.

**Please mark all sections**

| A. Clinical Skills | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| History Taking | | | | | |
| Physical Examination | | | | | |
| Diagnostic Investigations (Lab / Imaging) | | | | | |
| Diagnostic Skills | | | | | |
| Clinical Judgement | | | | | |
| Operative Skills | | | | | |
| Post-operative Management | | | | | |
| Follow-up Planning | | | | | |
| **B. Professional Development** | | | | | |
| Teaching Activities | | | | | |
| Clinical Audit | | | | | |
| Presentations | | | | | |
| Research | | | | | |

RCSI

| C. Personal Skills | | | | | |
|---|---|---|---|---|---|
| Communication | | | | | |
| Teamwork | | | | | |
| Leadership | | | | | |
| Self-Awareness and Insight | | | | | |
| Motivation and Drive | | | | | |
| Disposition and Appearance | | | | | |
| Management of Stress and Workload | | | | | |
| Management of Crises | | | | | |
| Reliability | | | | | |
| Time Management | | | | | |
| **D. Relationships** | | | | | |
| Medical Colleagues | | | | | |
| Nursing and Paramedical Staff | | | | | |
| Patients and Relatives | | | | | |

**Ratings 1 and 2 are deemed Below Expectations, and require further detail and proposed action to address.**

**Assessor's Signature: ……**

**Assessor's Name (Print): ……**

**Assessor's IMC/GMC number……**

**Date: ……**

## 8.4 Supplemental Material – Chapter 4

### 8.4.1 Appendix 4.1.1 – SSAOP tool

RCSI SURGICAL AFFAIRS

**NATIONAL SURGICAL TRAINING CENTRE**

**Supervised Structured Assessment of Operative Performance (SSAOP)**

RCSI DEVELOPING HEALTHCARE LEADERS WHO MAKE A DIFFERENCE WORLDWIDE

| | | |
|---|---|---|
| **Trainee Name:** | **Trainer Name:** | |
| **Hospital:** | **Specialty:** | |
| **Programme Year:** | **Rotation Start Date:** | **End Date:** |

**Procedure Performed:**

This is an assessment of operative performance of the surgical trainee. Trainees must complete three SSAOPs in each six month rotation (i.e. 12 in total during CST). Each SSAOP should be for a different surgical procedure. The assessment can be performed on any surgical procedure which the trainer considers appropriate for the specialty and the trainee year of training e.g. appendectomy, or repair of inguinal hernia. Minor operations (e.g. excision of skin lesions) are unacceptable.

**NB. The completed form will NOT be considered if "N/A" is entered for three or more of the items numbered 1 -15 (i.e. 20% of the items)**

**A maximum of 15 points is awarded per SSAOP. Each SSAOP is scored out of 150 for ease of calculation.**

**N/A = not observed or not done by the trainee**

**Please mark all sections by checking the box in the appropriate category.**

**Case Difficulty (Max 15 marks)**

| Straightforward anatomy, no related prior operation(s) or treatment (e.g. radiation) (5) | Intermediate difficulty (10) | Obesity, abnormal anatomy, complex pathology, related prior operation(s) or treatment (15) |
|---|---|---|
| ☐ | ☐ | ☐ |

**Degree of Prompting or Direction (Max 15 marks)**

| Substantial direction by consultant. Trainee performs all steps but the consultant provides constant direction to the trainee and surgical team. (5) | Some direction by consultant. Trainee performs all steps but the consultant provides occasional direction to the trainee and /or the surgical team. (10) | Minimal direction by consultant. Trainee performs all steps and directs the surgical team independently with little or no direction from the consultant to either the trainee or the surgical team. (15) |
|---|---|---|
| ☐ | ☐ | ☐ |

| **Pre-Operative Preparation (Max 30 marks)**<br>***Please indicate the degree to which trainee:*** | Very poor. Unacceptable for level of training<br>(1) | Below expectations for level of training<br>(2) | Meets expectations for level of training<br>(3) | Above expectations for level of training<br>(4) | Exceptional. Capable of performing independently<br>(5) | N/A |
|---|---|---|---|---|---|---|
| 1. Demonstrates knowledge of case & relevant investigations. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Ensures availability of relevant patient data & x-rays. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Ensures patient is appropriately positioned with attention to pressure areas & diathermy pad. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Initiates "Time out" in accordance with WHO guidelines | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5.Demonstrates effective scrubbing & gowning technique | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Demonstrates appropriate skin preparation & draping. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| **Operative Procedure (Max 80 marks).**<br>***Please indicate the degree to which trainee:*** | Very poor. Unacceptable for level of training<br>(2) | Below expectations for level of training<br>(4) | Meets expectations for level of training<br>(6) | Above expectations for level of training<br>(8) | Exceptional. Capable of performing independently<br>(10) | N/A |
|---|---|---|---|---|---|---|
| 7. Plans & makes incision(s) taking into account procedure & prior surgery. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. Chooses appropriate surgical instruments (including sutures). | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. Uses surgical instruments (includes diathermy, stapler etc.) in a manner that is safe (i.e. visualizes tips) & demonstrates economy of motion. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. Handles tissue carefully, demonstrates meticulous dissection technique & appropriate use of tissue planes. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 11. Anticipates needs, thinks ahead & sets up operative field by effective use of retraction & assistants, with efficient forward progression & flow. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 12. Demonstrates a calm demeanor, is not distracted, slows down when appropriate & accepts feedback. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | | | | | | |
|---|---|---|---|---|---|---|
| **13. Is able to problem solve, recognise, avoid & recover from errors & unexpected events (includes abnormal anatomy & bleeding).** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| **14. Communicates effectively with team, provides leadership & appropriately monitors changes outside immediate operative field.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| **Postoperative Care (10)** *Please indicate the degree to which the trainee:* | Very poor. Unacceptable for level of training (2) | Below expectations for level of training (4) | Meets expectations for level of training (6) | Above expectations for level of training (8) | Exceptional. Capable of performing independently (10) | N/A |
|---|---|---|---|---|---|---|
| **15. Completes the operation note & writes post-operative orders.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

**Please indicate the weaknesses in this trainee's performance:**

**Please indicate the strengths in this trainee's performance:**

Please rate this trainee's overall performance in this procedure. Exceptional rating indicates technically proficient performance (i.e., trainee is ready to perform **THIS** operation independently assuming he/she consistently performs at this level).

| **Unable to perform basic operative steps without direction** | **Fair performance but needed considerable direction to complete steps.** | **Good performance but needed some direction** | **Excellent performance but needed some direction** | **Exceptional performance. Could operate independently.** |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

Assessor's Signature: _____

Assessor's Name _____ (Print)

Date _____

Office Use Only

Score: _____

## 8.4.2 Appendix 4.2.1 – SSAOP tool

RCSI SURGICAL AFFAIRS

# NATIONAL SURGICAL TRAINING CENTRE

**Supervised Structured Assessment of Operative Performance (SSAOP)**

RCSI DEVELOPING HEALTHCARE LEADERS WHO MAKE A DIFFERENCE WORLDWIDE

| | | |
|---|---|---|
| **Trainee Name:** | **Trainer Name:** | |
| **Hospital:** | **Specialty:** | |
| **Programme Year:** | **Rotation Start Date:** | **End Date:** |

**Procedure Performed:**

This is an assessment of operative performance of the surgical trainee. Trainees must complete three SSAOPs in each six month rotation (i.e. 12 in total during CST). Each SSAOP should be for a different surgical procedure. The assessment can be performed on any surgical procedure which the trainer considers appropriate for the specialty and the trainee year of training e.g. appendectomy, or repair of inguinal hernia. Minor operations (e.g. excision of skin lesions) are unacceptable.

**NB. The completed form will NOT be considered if "N/A" is entered for three or more of the items numbered 1 -15 (i.e. 20% of the items)**

**A maximum of 15 points is awarded per SSAOP. Each SSAOP is scored out of 150 for ease of calculation.**

**N/A = not observed or not done by the trainee**

**Please mark all sections by checking the box in the appropriate category.**

**Case Difficulty (Max 15 marks)**

| Straightforward anatomy, no related prior operation(s) or treatment (e.g. radiation) (5) | Intermediate difficulty (10) | Obesity, abnormal anatomy, complex pathology, related prior operation(s) or treatment (15) |
|---|---|---|
| ☐ | ☐ | ☐ |

**Degree of Prompting or Direction (Max 15 marks)**

| Substantial direction by consultant. Trainee performs all steps but the consultant provides constant direction to the trainee and surgical team. (5) | Some direction by consultant. Trainee performs all steps but the consultant provides occasional direction to the trainee and /or the surgical team. (10) | Minimal direction by consultant. Trainee performs all steps and directs the surgical team independently with little or no direction from the consultant to either the trainee or the surgical team. (15) |
|---|---|---|
| ☐ | ☐ | ☐ |

| **Pre-Operative Preparation (Max 30 marks)**<br>***Please indicate the degree to which trainee:*** | Very poor. Unacceptable for level of training<br>(1) | Below expectations for level of training<br>(2) | Meets expectations for level of training<br>(3) | Above expectations for level of training<br>(4) | Exceptional. Capable of performing independently<br>(5) | N/A |
|---|---|---|---|---|---|---|
| 1. Demonstrates knowledge of case & relevant investigations. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Ensures availability of relevant patient data & x-rays. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Ensures patient is appropriately positioned with attention to pressure areas & diathermy pad. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Initiates "Time out" in accordance with WHO guidelines | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5.Demonstrates effective scrubbing & gowning technique | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Demonstrates appropriate skin preparation & draping. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| **Operative Procedure (Max 80 marks).**<br>***Please indicate the degree to which trainee:*** | Very poor. Unacceptable for level of training<br>(2) | Below expectations for level of training<br>(4) | Meets expectations for level of training<br>(6) | Above expectations for level of training<br>(8) | Exceptional. Capable of performing independently<br>(10) | N/A |
|---|---|---|---|---|---|---|
| 7. Plans & makes incision(s) taking into account procedure & prior surgery. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. Chooses appropriate surgical instruments (including sutures). | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. Uses surgical instruments (includes diathermy, stapler etc.) in a manner that is safe (i.e. visualizes tips) & demonstrates economy of motion. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. Handles tissue carefully, demonstrates meticulous dissection technique & appropriate use of tissue planes. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 11. Anticipates needs, thinks ahead & sets up operative field by effective use of retraction & assistants, with efficient forward progression & flow. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 12. Demonstrates a calm demeanor, is not distracted, slows down when appropriate & accepts feedback. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | | | | | | |
|---|---|---|---|---|---|---|
| **13. Is able to problem solve, recognise, avoid & recover from errors & unexpected events (includes abnormal anatomy & bleeding).** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| **14. Communicates effectively with team, provides leadership & appropriately monitors changes outside immediate operative field.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| **Postoperative Care (10)** *Please indicate the degree to which the trainee:* | Very poor. Unacceptable for level of training (2) | Below expectations for level of training (4) | Meets expectations for level of training (6) | Above expectations for level of training (8) | Exceptional. Capable of performing independently (10) | N/A |
|---|---|---|---|---|---|---|
| **15. Completes the operation note & writes post-operative orders.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

**Please indicate the weaknesses in this trainee's performance:**

**Please indicate the strengths in this trainee's performance:**

Please rate this trainee's overall performance in this procedure. Exceptional rating indicates technically proficient performance (i.e., trainee is ready to perform **THIS** operation independently assuming he/she consistently performs at this level).

| **Unable to perform basic operative steps without direction** | **Fair performance but needed considerable direction to complete steps.** | **Good performance but needed some direction** | **Excellent performance but needed some direction** | **Exceptional performance. Could operate independently.** |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

Assessor's Signature: _____

Assessor's Name _____ (Print)

Date _____

Office Use Only

Score: _____

### 8.4.3 Appendix 4.2.2

*Appendix 4.2.2: Operative/ technical competency expectations for core surgical trainees pursuing subspecialty training in general surgery.*

| Critical Skills (all trainees) – Operative competencies | Level of performance required |
|---|---|
| - effective hand washing/ gloving and gowning | 4 |
| - accurate, effective and safe administration of local anaesthetic | 3<br>3 |
| - preparation and maintenance of an aseptic field | 3 |
| - incision of skin and subcutaneous tissue | 3 |
| - closure of skin and subcutaneous tissue | 3 |
| - completion of the WHO checklist | |
| Core Specialty Modules – Operative General Surgery – Technical Skills and Procedures | |
| - Excision biopsy of skin lesion | 2 |
| - Repair of primary abdominal wall hernia | 1 |
| - Open and close midline laparotomy | 1 |
| - Placement of laparoscopic ports | 1 |
| - Appendicectomy | 1 |
| - Superficial abscess drainage | 2 |
| ST3 Preparation Modules – General Surgery – Technical Skills and Procedures | |
| - Chest drain insertion | 3 |
| - Needle biopsy including fine needle aspiration | 3 |
| - Rigid sigmoidoscopy | 3 |
| - Excision biopsy of a benign skin or subcutaneous lesion | 4 |
| - Outpatient treatment of haemorrhoids | 2 |
| - Induction of pneumoperitoneum for laparoscopy with port placement | 2 |
| - Open and close midline laparotomy incision | 2 |
| - Appendicectomy | 3 |
| - Inguinal hernia repair | 2 |
| - Primary abdominal wall hernia repair | 2 |

*Level of performance required defined as; 1 – Has observed, 2 – Can do with assistance, 3 – Can do whole but may need assistance, and 4 – competent to do without assistance, including complications. Adapted from the Intercollegiate Surgical Curriculum Program 'Core Surgical Training Curriculum 2021' (Brecknell J, Lintott P, 2021).*

### 8.4.4 Appendix 4.2.3 - Generalizability theory – methodology and model development.

Available at: Toale C, Morris M, O'Keeffe D, Boland F, Ryan DM, Nally DM, Kavanagh DO. Assessing operative competence in core surgical training: A reliability analysis. Am J Surg. 2023 Nov;226(5):588-595. doi: 10.1016/j.amjsurg.2023.06.020. Epub 2023 Jun 30. PMID: 37481408.

### 8.4.5 Appendix 4.2.4 - *Number of assessors and observations required to reliably assesses operative performance in Core Surgical Training.*

Available at: Toale C, Morris M, O'Keeffe D, Boland F, Ryan DM, Nally DM, Kavanagh DO. Assessing operative competence in core surgical training: A reliability analysis. Am J Surg. 2023 Nov;226(5):588-595. doi: 10.1016/j.amjsurg.2023.06.020. Epub 2023 Jun 30. PMID: 37481408.

## 8.4.6 Appendix 4.3.1 – Example Objective Surgical Skill (OSS) Tool

RCSI

### Operative Skills Score Sheet 2018

**Assessor Name (legible):** _____

**IMC number for CPD:** _____ **Candidate ID:**

**Date:** _____ **Time:** _____

Procedure: **BOWEL ANASTOMOSIS**

Instructions: **Perform an end to end bowel anastomosis using a single layer of interrupted inverting sutures, hand tied. You may request assistance if required.**

| | Not Done | Done |
|---|---|---|
| **Alignment of bowel ends** | | |
| 1.Candidate aligns bowel ends correctly (mesenteric to mesenteric side) | | |
| 2. Secures corners (three o'clock and nine o'clock) with stay sutures | | |
| | | |
| **Suture placement** | | |
| 4.Places adequate number of sutures (i.e., every 3mm approx; > every 5mm = incorrect) | | |
| 5. Sutures appropriate distance from bowel edge (ie, ~3 mm; >4mm = incorrect) | | |
| 6. Sutures placed at equal intervals and parallel to each other | | |
| **Knot tying** | | |
| 7. Hand-tied knots with minimum 3 throws | | |
| 8. All throws square | | |
| **Apposition of bowel** | | |
| 9. Sutures tied at correct tension | | |
| 10. Bowel evenly apposed (without dog ears i.e., puckered area of bowel between two stitches) | | |
| 11. Bowel edges inverted | | |
| 12. All knots on the inside (except final knot which may be external) | | |
| 12. Gap-free anastomosis ( likely to be water-tight in vivo) | | |
| 13. Gentle tissue handling throughout | | |
| **Total** | | |

| **What is the overall competence of this candidate? (independent of the scores above)** | | |
|---|---|---|
| **Not Competent/Fail** ☐ | **Borderline** ☐ | **Competent/Pass** ☐ |

**Comments:**

**PLEASE TURN OVER TO COMPLETE DETAILED GLOBAL RATING PAGE** ➡

*Updated April 2018*

## 8.4.7 Appendix 4.4.1 – SSAOP tool

RCSI SURGICAL AFFAIRS

# NATIONAL SURGICAL TRAINING CENTRE

**Supervised Structured Assessment of Operative Performance (SSAOP)**

RCSI DEVELOPING HEALTHCARE LEADERS WHO MAKE A DIFFERENCE WORLDWIDE

| | | |
|---|---|---|
| **Trainee Name:** | **Trainer Name:** | |
| **Hospital:** | **Specialty:** | |
| **Programme Year:** | **Rotation Start Date:** | **End Date:** |

**Procedure Performed:**

This is an assessment of operative performance of the surgical trainee. Trainees must complete three SSAOPs in each six month rotation (i.e. 12 in total during CST). Each SSAOP should be for a different surgical procedure. The assessment can be performed on any surgical procedure which the trainer considers appropriate for the specialty and the trainee year of training e.g. appendectomy, or repair of inguinal hernia. Minor operations (e.g. excision of skin lesions) are unacceptable.

**NB. The completed form will NOT be considered if "N/A" is entered for three or more of the items numbered 1 -15 (i.e. 20% of the items)**

**A maximum of 15 points is awarded per SSAOP. Each SSAOP is scored out of 150 for ease of calculation.**

**N/A = not observed or not done by the trainee**

**Please mark all sections by checking the box in the appropriate category.**

**Case Difficulty (Max 15 marks)**

| | | |
|---|---|---|
| Straightforward anatomy, no related prior operation(s) or treatment (e.g. radiation)<br>(5) | Intermediate difficulty<br>(10) | Obesity, abnormal anatomy, complex pathology, related prior operation(s) or treatment<br>(15) |
| ☐ | ☐ | ☐ |

**Degree of Prompting or Direction (Max 15 marks)**

| | | |
|---|---|---|
| Substantial direction by consultant. Trainee performs all steps but the consultant provides constant direction to the trainee and surgical team.<br>(5) | Some direction by consultant. Trainee performs all steps but the consultant provides occasional direction to the trainee and /or the surgical team.<br>(10) | Minimal direction by consultant. Trainee performs all steps and directs the surgical team independently with little or no direction from the consultant to either the trainee or the surgical team.<br>(15) |
| ☐ | ☐ | ☐ |

| **Pre-Operative Preparation (Max 30 marks)**<br>***Please indicate the degree to which trainee:*** | Very poor. Unacceptable for level of training<br>(1) | Below expectations for level of training<br>(2) | Meets expectations for level of training<br>(3) | Above expectations for level of training<br>(4) | Exceptional. Capable of performing independently<br>(5) | N/A |
|---|---|---|---|---|---|---|
| 1. Demonstrates knowledge of case & relevant investigations. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. Ensures availability of relevant patient data & x-rays. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. Ensures patient is appropriately positioned with attention to pressure areas & diathermy pad. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. Initiates "Time out" in accordance with WHO guidelines | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5.Demonstrates effective scrubbing & gowning technique | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. Demonstrates appropriate skin preparation & draping. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| **Operative Procedure (Max 80 marks).**<br>***Please indicate the degree to which trainee:*** | Very poor. Unacceptable for level of training<br>(2) | Below expectations for level of training<br>(4) | Meets expectations for level of training<br>(6) | Above expectations for level of training<br>(8) | Exceptional. Capable of performing independently<br>(10) | N/A |
|---|---|---|---|---|---|---|
| 7. Plans & makes incision(s) taking into account procedure & prior surgery. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. Chooses appropriate surgical instruments (including sutures). | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. Uses surgical instruments (includes diathermy, stapler etc.) in a manner that is safe (i.e. visualizes tips) & demonstrates economy of motion. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. Handles tissue carefully, demonstrates meticulous dissection technique & appropriate use of tissue planes. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 11. Anticipates needs, thinks ahead & sets up operative field by effective use of retraction & assistants, with efficient forward progression & flow. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 12. Demonstrates a calm demeanor, is not distracted, slows down when appropriate & accepts feedback. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | | | | | | |
|---|---|---|---|---|---|---|
| **13. Is able to problem solve, recognise, avoid & recover from errors & unexpected events (includes abnormal anatomy & bleeding).** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| **14. Communicates effectively with team, provides leadership & appropriately monitors changes outside immediate operative field.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| **Postoperative Care (10)** *Please indicate the degree to which the trainee:* | Very poor. Unacceptable for level of training (2) | Below expectations for level of training (4) | Meets expectations for level of training (6) | Above expectations for level of training (8) | Exceptional. Capable of performing independently (10) | N/A |
|---|---|---|---|---|---|---|
| **15. Completes the operation note & writes post-operative orders.** | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

**Please indicate the weaknesses in this trainee's performance:**

**Please indicate the strengths in this trainee's performance:**

Please rate this trainee's overall performance in this procedure. Exceptional rating indicates technically proficient performance (i.e., trainee is ready to perform **THIS** operation independently assuming he/she consistently performs at this level).

| **Unable to perform basic operative steps without direction** | **Fair performance but needed considerable direction to complete steps.** | **Good performance but needed some direction** | **Excellent performance but needed some direction** | **Exceptional performance. Could operate independently.** |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

Assessor's Signature: _____

Assessor's Name _____ (Print)

Date _____

Office Use Only

Score: _____

## 8.5 Supplemental Data – Chapter 5

### 8.5.1 Appendix 5.1.1

Available at: Toale C, Morris M, Konge L, Nayahangan LJ, Roche A, Heskin L, Kavanagh DO. Generating a Prioritised List of Operative Procedures for Simulation-Based Assessment of General Surgery Trainees through Consensus. Ann Surg. 2023 Oct 9. doi: 10.1097/SLA.0000000000006118. Epub ahead of print. PMID: 37811854.

### 8.5.2 Appendix 5.1.2 - Delphi Consensus Model, Survey 1

Available at: Toale C, Morris M, Konge L, Nayahangan LJ, Roche A, Heskin L, Kavanagh DO. Generating a Prioritised List of Operative Procedures for Simulation-Based Assessment of General Surgery Trainees through Consensus. Ann Surg. 2023 Oct 9. doi: 10.1097/SLA.0000000000006118. Epub ahead of print. PMID: 37811854.

### 8.5.3 Appendix 5.2.1

Data extraction pro-forma

Available at: Toale C, Morris M, Kavanagh DO. Training and assessment using the LapSim laparoscopic simulator: a scoping review of validity evidence. Surg Endosc. 2023 Mar;37(3):1658-1671. doi: 10.1007/s00464-022-09593-0. Epub 2022 Sep 19. PMID: 36123545.

### 8.5.4 Appendix 5.2.2 – Medical Educational Research Study Quality Instrument (MERSQI) scores for included studies

Toale C, Morris M, Kavanagh DO. Training and assessment using the LapSim Available at: laparoscopic simulator: a scoping review of validity evidence. Surg Endosc. 2023 Mar;37(3):1658-1671. doi: 10.1007/s00464-022-09593-0. Epub 2022 Sep 19. PMID: 36123545.

### 8.5.5

## Appendix 5.3.1 – LapSim™ recorded metrics

---

Laparoscopic Cholecystectomy:

- Left instrument path length (m)
- Left instrument angular path (degrees)
- Right instrument path length (m)
- Right instrument angular path (degrees)
- Left instrument outside view (#)
- Left instrument outside view (s)
- Right instrument outside view (#)
- Right instrument outside view (s)
- Blood loss (ml)
- Energy applied in air (s)
- Missing clips (#)
- Fatal clips (#)
- Missing cuts (#)
- Number of cuts on the liver (#)
- Number of cuts on the gallbladder (#)
- Ducts/ vessels ripped (#)
- Ducts/ vessels ablated (#)
- Number of fatal dissections (#)
- Total time burned on arteries (s)
- Total time burned on ducts (s)
- Total time burned on liver (s)
- Number of times burned on arteries (#)
- Number of times burned on ducts (#)
- Number of times burned on liver (#)
- Number of times burned on gallbladder (#)
- Percent of connective tissue ablated
- Percent of adipose tissue ablated
- Percent of connective tissue ripped
- Percent of adipose tissue ripped

---

Laparoscopic appendicectomy

- Total time
- Left instrument path length (m)
- Left instrument angular path (degrees)
- Right instrument path length (m)
- Right instrument angular path (degrees)
- Left instrument outside view (#)
- Left instrument outside view (s)
- Right instrument outside view (#)
- Right instrument outside view (s)
- Blood loss (ml)
- Energy applied in air (s)
- Total time burned on the appendix (s)
- Number of times burned on the appendix (#)
- Percent of adhesions ablated
- Percent of mesentery ablated
- Percent of adhesions ripped

- Percent of mesentery ripped
- Evacuation from body (performed/ not performed)
- Incorrect stapler used (yes/ no)

---

Laparoscopic right hemi-colectomy

- Total time
- Left instrument path length (m)
- Left instrument angular path (degrees)
- Left instrument outside view (#)
- Left instrument outside view (s)
- Right instrument path length (m)
- Right instrument angular path (degrees)
- Right instrument outside view (#)
- Right instrument outside view (s)
- Number of vessel injured (#)
- Correct vessel divisions (#)
- Burn time on artery (s)
- Burn time on vein (s)
- Burn time on duodenum (s)
- Number of burns on artery (#)
- Number of burns on vein (#)
- Number of burns on duodenum (#)
- Percentage of artery ablated
- Percentage of vein ablated
- Total time artery overstretched (s)
- Total time vein overstretched (s)
- Number of times artery overstretched
- Number of veins overstretched
- Artery ripped (yes/no)
- Vein ripped (yes/no)
- Number of pressure damages on duodenum (#)
- Gonadal artery burned (yes/ no)
- Burned on ureter (yes/ no)
- Maryland jaw vessel sealer/divider has been used on wrong structures (yes/no)

---

### 8.5.6 Appendix 5.3.2

## Internal consistency

*Internal consistency of assessment tools, assessed using Cronbach's alpha*

| **Score** | □ |
|---|---|
| PBA | 0.68 |
| OSATS | 0.85 |
| Total Score | 0.81 |
| PBA global rater | 0.81 |
| Competency rater | 0.71 |

*PBA indicates Procedure-Based Assessment tool, OSATS, Objective Structured Assessment of Technical Skill.*

### 8.5.6 Appendix 5.3.3 – Correlations between individual station and total scores

*Correlation (Pearson) between individual station scores and total scores awarded*

| Station | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| **r** | .48 | .66 | .63 | .69 | .54 | .89 | .84 | .57 |
| **p** | .10 | .02 * | .02 * | .01 * | .07 * | <.001 *** | <.001 *** | .06 |

*r indicates Pearson correlation coefficient, p, p-value, *, p<0.05, **, p<0.005, ***, p<0.0005*

## 8.5.7 Appendix 5.3.4

Correlations between score domains

*Internal consistency: correlations between score domains at the individual station level*

| Station 1 – fistula-in-ano | | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1. mPBA | | | | |
| 2. OSATS | **τb = .83 p <0.001** | | | |
| 3. Total | **τb = .94 p <.0001** | **τb = .90 p <.0001** | | |
| 4. mPBA GR | **τb = .77 p = .001** | **τb = .83 p <.0001** | **τb = .79 p = 0.001** | |
| 5. Competence | **τb = .63 p = .010** | **τb = .72 p = .003** | **τb = .67 p = 0.005** | **τb = .59 p = 0.024** |

| Station 2 - ileostomy closure | | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1. mPBA | | | | |
| 2. OSATS | **τb = .68 p = 0.003** | | | |
| 3. Total | **τb = .87 p <.0001** | **τb = .83 p <.0001** | | |
| 4. mPBA GR | **τb = .61 p = 0.009** | **τb = .66 p = 0.005** | **τb = .52 p = .022** | |
| 5. Competence | **τb = .51 p = 0.040** | **τb = .57 p = 0.021** | **τb = .49 p = 0.05** | **τb = .68 p = 0.007** |

| Station 3 - right hemi-colectomy (vessel ligation) | | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1. mPBA | | | | |
| 2. OSATS | **τb = .73 p = 0.001** | | | |
| 3. Total | **τb = .90 p<.0001** | **τb = .84 p <.0001** | | |
| 4. mPBA GR | **τb = .79 p.= .001** | **τb = .73 p – 0.001** | **τb = .81 p<.0001** | |
| 5. Competence | **τb = .65 p = .010** | **τb = .53 p .033** | **τb = .62 p = .011** | **τb = .62 p = .019** |

| Station 4 - pilonidal sinus excision | | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1. mPBA | | | | |
| 2. OSATS | τb = 0.013 p = .951 | | | |
| 3. Total | **τb = .56 p = 0.007** | **τb = .46 p = .032** | | |
| 4. mPBA GR | τb = -.11 p = .647 | **τb = .80 p = .001** | τb = .33 p = .151 | |
| 5. Competence | **τb = -.099 p = 0.68** | **τb = .72 p = .003** | τb = .36 p = .131 | **τb = .89 p = .001** |

| Station 5 - ventral hernia repair | | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1. mPBA | | | | |
| 2. OSATS | **τb = .69 p = .003** | | | |
| 3. Total | **τb = .95 p<.0001** | **τb = .75 p = .001** | | |
| 4. mPBA GR | **τb = .73 p = .003** | **τb = .69 p = .006** | **τb = .77 p = .001** | |
| 5. Competence | **τb = .72 p = .004** | **τb = .79 p = .002** | **τb =.73 p = .003** | **τb = .74 p = .006** |

| Station 6 laparoscopic appendicectomy | | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1. mPBA | | | | |
| 2. OSATS | τb = .58 p = .011 | | | |
| 3. Total | **τb = .79 p<.0001** | **τb = .82 p<.0001** | | |
| 4. mPBA GR | **τb = .64 p = .007** | **τb = .73 p = .001** | **τb = .70 p = .002** | |
| 5. Competence | τb = .337 p = .186 | **τb = .62 p =.011** | τb = .45 p = .065 | **τb = .54 p = .032** |

| Station 7 – emergency laparotomy (blunt liver trauma) | | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1. Checklist | | | | |
| 2. OSATS | **τb = .88 p<.0001** | | | |
| 3. Total | **τb = .97 p<.0001** | **τb = .91 p<.0001** | | |
| 4. mPBA GR | **τb = . 88 p<.0001** | **τb = .89 p<.0001** | **τb = .89 p<.0001** | |
| 5. Competence | **τb = .74 p = .002** | **τb = .74 p = .002** | **τb = .74 p = .002** | **τb = .77 p = .003** |

| Station 8 - laparoscopic cholecystectomy | | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1. Checklist | | | | |
| 2. OSATS | **τb = .56 p = .026** | | | |
| 3. Total | **τb = .72 p =.004** | **τb = .87 p<.001** | | |
| 4. mPBA GR | **τb = .74 p = .004** | **τb = .53 p =.038** | **τb = .58 p = .020** | |
| 5. Competence | τb = .38 p = .163 | τb = .47 p = .088 | τb = .44 p = .099 | τb = .39 p = .161 |

*mPBA indicates modified Procedure Based Assessment (checklist) score, OSATS, Objective Structured Assessment of Technical Skills, mPBA GR, mPBA Global Rater, Competence, global competence rater, τb, Kendalls tau-b, p, p-value. Significant correlations are highlighted in bold text.*

## 8.5.8 Appendix 5.3.5

*Automated LapSim performance metrics for junior vs senior trainees; difference mean results for laparoscopic cholecystectomy by group, junior vs senior trainees*

**Independent Samples Test**

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 95% Confidence Interval of the Difference | |
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| Total time | Equal variances assumed | .842 | .380 | .885 | 10 | .397 | 80.08112 | 90.52936 | -121.63087 | 281.79310 |
| | Equal variances not assumed | | | .851 | 7.518 | .421 | 80.08112 | 94.07780 | -139.30572 | 299.46796 |
| Left instrument path length | Equal variances assumed | .121 | .735 | .828 | 10 | .427 | .93245 | 1.12648 | -1.57750 | 3.44241 |
| | Equal variances not assumed | | | .819 | 8.432 | .435 | .93245 | 1.13853 | -1.66976 | 3.53467 |
| Left instrument angular path | Equal variances assumed | .000 | .991 | .315 | 10 | .759 | 109.72449 | 348.52849 | -666.84538 | 886.29436 |
| | Equal variances not assumed | | | .324 | 9.579 | .753 | 109.72449 | 338.38591 | -648.75902 | 868.20800 |
| Right instrument path length | Equal variances assumed | .544 | .478 | .358 | 10 | .728 | .60225 | 1.68198 | -3.14544 | 4.34994 |
| | Equal variances not assumed | | | .350 | 8.015 | .735 | .60225 | 1.72168 | -3.36664 | 4.57114 |
| Right instrument angular path | Equal variances assumed | 8.162 | .017 | 1.321 | 10 | .216 | 462.48162 | 350.20296 | -317.81921 | 1242.78244 |
| | Equal variances not assumed | | | 1.184 | 5.504 | .285 | 462.48162 | 390.48463 | -514.25398 | 1439.21721 |
| Left instrument outside view (n) | Equal variances assumed | 1.208 | .297 | .808 | 10 | .438 | 3.91429 | 4.84348 | -6.87765 | 14.70622 |
| | Equal variances not assumed | | | .743 | 6.151 | .485 | 3.91429 | 5.26604 | -8.89514 | 16.72371 |
| Left instrument outside view (s) | Equal variances assumed | .389 | .547 | -1.072 | 10 | .309 | -108.84718 | 101.50742 | -335.01979 | 117.32544 |
| | Equal variances not assumed | | | -1.082 | 9.040 | .307 | -108.84718 | 100.61127 | -336.29082 | 118.59647 |
| Right instrument outside view (n) | Equal variances assumed | 2.702 | .131 | .751 | 10 | .470 | 5.28571 | 7.03971 | -10.39974 | 20.97117 |
| | Equal variances not assumed | | | .706 | 6.790 | .504 | 5.28571 | 7.48654 | -12.52892 | 23.10035 |
| Right instrument outside view (s) | Equal variances assumed | 4.561 | .058 | 1.161 | 10 | .273 | 12.90408 | 11.11436 | -11.86025 | 37.66842 |
| | Equal variances not assumed | | | 1.051 | 5.737 | .335 | 12.90408 | 12.27568 | -17.47054 | 43.27870 |
| Blood loss | Equal variances assumed | 6.226 | .032 | 4.266 | 10 | **.002** | .79746 | .18693 | .38095 | 1.21397 |
| | Equal variances not assumed | | | 3.863 | 5.738 | .009 | .79746 | .20646 | .28663 | 1.30829 |
| Energy applied in air | Equal variances assumed | .778 | .398 | .074 | 10 | .942 | .32297 | 4.33932 | -9.34565 | 9.99158 |
| | Equal variances not assumed | | | .080 | 9.997 | .938 | .32297 | 4.05576 | -8.71421 | 9.36014 |
| Missing clips | Equal variances assumed | 2.538 | .142 | -.750 | 10 | .470 | -.48571 | .64751 | -1.92845 | .95702 |
| | Equal variances not assumed | | | -.869 | 7.646 | .411 | -.48571 | .55867 | -1.78446 | .81304 |
| Fatal clips | Equal variances assumed | 10.370 | .009 | 1.208 | 10 | .255 | .20000 | .16562 | -.16901 | .56901 |
| | Equal variances not assumed | | | 1.000 | 4.000 | .374 | .20000 | .20000 | -.35529 | .75529 |
| Number of cuts on the liver | Equal variances assumed | 4.000 | .073 | -.833 | 10 | .424 | -.14286 | .17143 | -.52482 | .23911 |
| | Equal variances not assumed | | | -1.000 | 6.000 | .356 | -.14286 | .14286 | -.49242 | .20670 |
| Number of cuts on the gallbladder | Equal variances assumed | .866 | .374 | -.564 | 10 | .585 | -.25714 | .45571 | -1.27253 | .75825 |
| | Equal variances not assumed | | | -.538 | 7.251 | .606 | -.25714 | .47752 | -1.37844 | .86415 |
| Total time burned on arteries | Equal variances assumed | 4.000 | .073 | -.833 | 10 | .424 | -.03185 | .03821 | -.11699 | .05330 |
| | Equal variances not assumed | | | -1.000 | 6.000 | .356 | -.03185 | .03185 | -.10977 | .04608 |
| Total time burned on liver | Equal variances assumed | .013 | .910 | .526 | 10 | .610 | 6.36167 | 12.08684 | -20.56948 | 33.29282 |
| | Equal variances not assumed | | | .522 | 8.517 | .615 | 6.36167 | 12.18402 | -21.44044 | 34.16378 |
| Total time burned on gallbladder | Equal variances assumed | .045 | .836 | 1.741 | 10 | .112 | 5.83055 | 3.34871 | -1.63084 | 13.29193 |
| | Equal variances not assumed | | | 1.807 | 9.731 | .102 | 5.83055 | 3.22707 | -1.38688 | 13.04798 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Number of times burned on arteries | Equal variances assumed | 4.000 | .073 | -.833 | 10 | .424 | -.14286 | .17143 | -.52482 | .23911 |
| | Equal variances not assumed | | | -1.000 | 6.000 | .356 | -.14286 | .14286 | -.49242 | .20670 |
| Number of times burned on liver | Equal variances assumed | .009 | .924 | .061 | 10 | .953 | .71429 | 11.69532 | -25.34450 | 26.77307 |
| | Equal variances not assumed | | | .063 | 9.593 | .951 | .71429 | 11.34778 | -24.71624 | 26.14481 |
| Number of times burned on gallbladder | Equal variances assumed | .531 | .483 | 3.022 | 10 | **.013** | 17.65714 | 5.84267 | 4.63886 | 30.67543 |
| | Equal variances not assumed | | | 3.195 | 9.970 | **.010** | 17.65714 | 5.52599 | 5.33938 | 29.97490 |
| Percentage of connective tissue ablated | Equal variances assumed | 3.135 | .107 | 1.275 | 10 | .231 | 2.33577 | 1.83215 | -1.74652 | 6.41806 |
| | Equal variances not assumed | | | 1.506 | 6.741 | .177 | 2.33577 | 1.55109 | -1.36078 | 6.03231 |
| Percentage of adipose tissue ablated | Equal variances assumed | 5.119 | .047 | -.079 | 10 | .939 | -.15535 | 1.96412 | -4.53169 | 4.22098 |
| | Equal variances not assumed | | | -.072 | 6.015 | .945 | -.15535 | 2.14618 | -5.40361 | 5.09291 |
| Percentage of connective tissue clipped | Equal variances assumed | 6.100 | .033 | .933 | 10 | .373 | .95516 | 1.02364 | -1.32566 | 3.23598 |
| | Equal variances not assumed | | | .781 | 4.164 | .477 | .95516 | 1.22281 | -2.38794 | 4.29827 |
| Percentage of adipose tissue clipped | Equal variances assumed | 1.563 | .240 | 4.336 | 10 | **.001** | 7.00946 | 1.61665 | 3.40734 | 10.61157 |
| | Equal variances not assumed | | | 4.096 | 6.935 | **.005** | 7.00946 | 1.71111 | 2.95567 | 11.06324 |

## 8.5.9 Appendix 5.3.6

Automated LapSim performance metrics for junior vs senior trainees; difference mean results for laparoscopic appendicectomy by group, junior vs senior trainees

**Independent Samples Test**

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 95% Confidence Interval of the Difference | |
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| Total time | Equal variances assumed | 6.824 | .024 | 2.843 | 11 | **.016** | 152.72931 | 53.71873 | 34.49519 | 270.96343 |
| | Equal variances not assumed | | | 2.662 | 6.072 | **.037** | 152.72931 | 57.36666 | 12.75991 | 292.69870 |
| Left instrument path length | Equal variances assumed | 1.110 | .315 | 1.186 | 11 | .261 | 1.38997 | 1.17221 | -1.19004 | 3.96999 |
| | Equal variances not assumed | | | 1.123 | 6.849 | .299 | 1.38997 | 1.23737 | -1.54904 | 4.32898 |
| Left instrument angular path | Equal variances assumed | 1.004 | .338 | .506 | 11 | .623 | 162.17713 | 320.39177 | -543.00041 | 867.35467 |
| | Equal variances not assumed | | | .479 | 6.815 | .647 | 162.17713 | 338.36719 | -642.36471 | 966.71897 |
| Right instrument path length | Equal variances assumed | .822 | .384 | .978 | 11 | .349 | 1.88256 | 1.92586 | -2.35623 | 6.12134 |
| | Equal variances not assumed | | | .928 | 7.013 | .384 | 1.88256 | 2.02828 | -2.91182 | 6.67693 |
| Right instrument angular path | Equal variances assumed | 1.181 | .300 | 1.547 | 11 | .150 | 681.83983 | 440.68782 | -288.10752 | 1651.78717 |
| | Equal variances not assumed | | | 1.476 | 7.340 | .182 | 681.83983 | 462.06278 | -400.57659 | 1764.25624 |
| Left instrument outside view (n) | Equal variances assumed | .535 | .480 | -2.184 | 11 | .051 | -6.78571 | 3.10637 | -13.62278 | .05135 |
| | Equal variances not assumed | | | -2.131 | 9.100 | .062 | -6.78571 | 3.18452 | -13.97750 | .40607 |
| Left instrument outside view (s) | Equal variances assumed | .159 | .698 | -3.210 | 11 | **.008** | -182.74900 | 56.92321 | -308.03613 | -57.46186 |
| | Equal variances not assumed | | | -3.128 | 9.010 | **.012** | -182.74900 | 58.42290 | -314.88836 | -50.60963 |
| Right instrument outside view (n) | Equal variances assumed | .889 | .366 | .012 | 11 | .991 | .07143 | 6.04166 | -13.22618 | 13.36904 |
| | Equal variances not assumed | | | .012 | 9.477 | .991 | .07143 | 6.16328 | -13.76473 | 13.90759 |
| Right instrument outside view (n | Equal variances assumed | .118 | .737 | .437 | 11 | .671 | 3.96295 | 9.07572 | -16.01257 | 23.93848 |
| | Equal variances not assumed | | | .440 | 10.955 | .668 | 3.96295 | 8.99703 | -15.84930 | 23.77521 |
| Blood loss | Equal variances assumed | 2.991 | .112 | .840 | 11 | .419 | 37.69252 | 44.89609 | -61.12311 | 136.50815 |
| | Equal variances not assumed | | | .813 | 8.515 | .438 | 37.69252 | 46.36822 | -68.11720 | 143.50224 |
| Energy applied in air | Equal variances assumed | .179 | .680 | .012 | 11 | .991 | .09691 | 7.99435 | -17.49854 | 17.69236 |
| | Equal variances not assumed | | | .012 | 9.496 | .991 | .09691 | 8.15313 | -18.20087 | 18.39470 |
| Total time burned on artery | Equal variances assumed | 3.330 | .095 | -.566 | 11 | .583 | -1.62376 | 2.86820 | -7.93662 | 4.68910 |
| | Equal variances not assumed | | | -.541 | 7.477 | .604 | -1.62376 | 3.00185 | -8.63119 | 5.38367 |
| Number of times burned on artery | Equal variances assumed | .884 | .367 | -3.174 | 11 | **.009** | -8.95238 | 2.82073 | -15.16076 | -2.74400 |
| | Equal variances not assumed | | | -3.255 | 10.789 | **.008** | -8.95238 | 2.75030 | -15.02026 | -2.88450 |
| Percentage of adipose tissue ablated | Equal variances assumed | .040 | .846 | -.031 | 11 | .976 | -.18229 | 5.80371 | -12.95616 | 12.59158 |
| | Equal variances not assumed | | | -.031 | 10.627 | .976 | -.18229 | 5.81287 | -13.03130 | 12.66672 |
| % of mesentery ablated | Equal variances assumed | .238 | .635 | .275 | 11 | .788 | .37403 | 1.35950 | -2.61821 | 3.36626 |
| | Equal variances not assumed | | | .280 | 10.993 | .785 | .37403 | 1.33780 | -2.57069 | 3.31874 |
| Percentage of adipose tissue ripped | Equal variances assumed | .509 | .491 | -.688 | 11 | .506 | -1.09747 | 1.59467 | -4.60731 | 2.41237 |
| | Equal variances not assumed | | | -.676 | 9.622 | .515 | -1.09747 | 1.62359 | -4.73443 | 2.53949 |
| Percentage of mesentery ripped | Equal variances assumed | 12.942 | .004 | 2.769 | 11 | **.018** | .38646 | .13955 | .07931 | .69361 |
| | Equal variances not assumed | | | 2.580 | 5.765 | **.043** | .38646 | .14978 | .01632 | .75660 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Appendix evacuated from body | Equal variances assumed | .106 | .751 | -.238 | 11 | .817 | -.07143 | .30074 | -.73336 | .59050 |
| | Equal variances not assumed | | | -.237 | 10.606 | .817 | -.07143 | .30136 | -.73773 | .59488 |
| Incorrect stapler used | Equal variances assumed | 4.874 | .049 | -.920 | 11 | .377 | -.14286 | .15530 | -.48467 | .19896 |
| | Equal variances not assumed | | | -1.000 | 6.000 | .356 | -.14286 | .14286 | -.49242 | .20670 |

## 8.5.10 Appendix 5.3.7

Automated LapSim performance metrics for junior vs senior trainees; difference mean results for laparoscopic right hemicolectomy by group, junior vs senior trainees

**Independent Samples Test**

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 95% Confidence Interval of the Difference | |
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| Total time | Equal variances assumed | .813 | .387 | -.807 | 11 | .437 | -123.58507 | 153.16246 | -460.69337 | 213.52322 |
| | Equal variances not assumed | | | -.837 | 10.196 | .422 | -123.58507 | 147.68473 | -451.79289 | 204.62275 |
| Left instrument path length | Equal variances assumed | .002 | .966 | -.359 | 11 | .726 | -.38172 | 1.06284 | -2.72102 | 1.95757 |
| | Equal variances not assumed | | | -.353 | 9.747 | .731 | -.38172 | 1.08024 | -2.79715 | 2.03370 |
| Left instrument angular path | Equal variances assumed | .026 | .874 | -.207 | 11 | .840 | -48.78813 | 235.96718 | -568.14839 | 470.57212 |
| | Equal variances not assumed | | | -.204 | 9.985 | .842 | -48.78813 | 239.00692 | -581.43831 | 483.86204 |
| Left instrument outside view (n) | Equal variances assumed | .030 | .867 | -.622 | 11 | .547 | -1.19048 | 1.91399 | -5.40315 | 3.02219 |
| | Equal variances not assumed | | | -.622 | 10.688 | .547 | -1.19048 | 1.91438 | -5.41907 | 3.03812 |
| Left instrument outside view (s) | Equal variances assumed | 2.331 | .155 | -.857 | 11 | .410 | -29.38232 | 34.27610 | -104.82351 | 46.05887 |
| | Equal variances not assumed | | | -.914 | 7.743 | .388 | -29.38232 | 32.13955 | -103.92616 | 45.16152 |
| Right instrument path length | Equal variances assumed | 8.860 | .013 | -.491 | 11 | .633 | -1.03952 | 2.11857 | -5.70246 | 3.62342 |
| | Equal variances not assumed | | | -.520 | 8.280 | .616 | -1.03952 | 1.99766 | -5.61920 | 3.54016 |
| Right instrument angular path | Equal variances assumed | 4.180 | .066 | -.062 | 11 | .952 | -23.69296 | 384.85679 | -870.75705 | 823.37113 |
| | Equal variances not assumed | | | -.065 | 9.370 | .950 | -23.69296 | 367.20381 | -849.39171 | 802.00578 |
| Right instrument outside view (n) | Equal variances assumed | .045 | .836 | 1.025 | 11 | .327 | 1.71429 | 1.67265 | -1.96720 | 5.39577 |
| | Equal variances not assumed | | | 1.034 | 10.959 | .323 | 1.71429 | 1.65780 | -1.93619 | 5.36476 |
| Right instrument outside view (s) | Equal variances assumed | 1.141 | .308 | -.087 | 11 | .932 | -.51386 | 5.92887 | -13.56321 | 12.53549 |
| | Equal variances not assumed | | | -.091 | 8.728 | .929 | -.51386 | 5.61701 | -13.28108 | 12.25336 |
| Number of vessels injured | Equal variances assumed | 1.496 | .247 | .189 | 11 | .854 | .09524 | .50457 | -1.01530 | 1.20578 |
| | Equal variances not assumed | | | .185 | 9.295 | .858 | .09524 | .51596 | -1.06632 | 1.25679 |
| Correct vessels divided | Equal variances assumed | 5.726 | .036 | -1.525 | 11 | .155 | -.52381 | .34339 | -1.27960 | .23198 |
| | Equal variances not assumed | | | -1.444 | 6.814 | .193 | -.52381 | .36266 | -1.38613 | .33851 |
| Burn time on artery | Equal variances assumed | 9.031 | .012 | 1.938 | 11 | .079 | 4.00472 | 2.06645 | -.54351 | 8.55295 |
| | Equal variances not assumed | | | 1.784 | 5.098 | .133 | 4.00472 | 2.24477 | -1.73240 | 9.74184 |
| Burn time on vein | Equal variances assumed | 3.302 | .097 | 1.067 | 11 | .309 | 1.10663 | 1.03725 | -1.17634 | 3.38960 |
| | Equal variances not assumed | | | .989 | 5.494 | .364 | 1.10663 | 1.11849 | -1.69254 | 3.90580 |
| Number of burns on artery | Equal variances assumed | 4.026 | .070 | -.765 | 11 | .460 | -10.90476 | 14.24854 | -42.26558 | 20.45606 |
| | Equal variances not assumed | | | -.831 | 6.083 | .437 | -10.90476 | 13.11958 | -42.90182 | 21.09229 |
| Number of burns on vein | Equal variances assumed | 4.307 | .062 | -.849 | 11 | .414 | -12.09524 | 14.25051 | -43.46040 | 19.26993 |
| | Equal variances not assumed | | | -.922 | 6.026 | .392 | -12.09524 | 13.11268 | -44.14675 | 19.95627 |
| Percentage of vein ablated | Equal variances assumed | 4.874 | .049 | -.920 | 11 | .377 | -.30857 | .33545 | -1.04690 | .42975 |
| | Equal variances not assumed | | | -1.000 | 6.000 | .356 | -.30857 | .30857 | -1.06362 | .44648 |
| Total time artery overstretched | Equal variances assumed | 24.275 | .000 | 2.031 | 11 | .067 | 2.94610 | 1.45046 | -.24634 | 6.13853 |
| | Equal variances not assumed | | | 1.873 | 5.197 | .118 | 2.94610 | 1.57262 | -1.05081 | 6.94301 |
| Total time vein overstretched | Equal variances assumed | .569 | .466 | .431 | 11 | .675 | .03626 | .08413 | -.14891 | .22143 |
| | Equal variances not assumed | | | .418 | 8.664 | .686 | .03626 | .08673 | -.16110 | .23362 |
| | Equal variances assumed | 1.321 | .275 | 1.025 | 11 | .327 | .88095 | .85963 | -1.01109 | 2.77299 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Number of times artery overstretched | Equal variances not assumed | | | .994 | 8.648 | .347 | .88095 | .88634 | -1.13661 | 2.89852 |
| Number of times vein overstretched | Equal variances assumed | 2.643 | .132 | -.732 | 11 | .480 | -.92857 | 1.26883 | -3.72124 | 1.86410 |
| | Equal variances not assumed | | | -.786 | 7.074 | .457 | -.92857 | 1.18135 | -3.71610 | 1.85896 |
| Vessels ripped | Equal variances assumed | .114 | .742 | .171 | 11 | .867 | .04762 | .27878 | -.56596 | .66120 |
| | Equal variances not assumed | | | .170 | 10.471 | .868 | .04762 | .28010 | -.57271 | .66795 |
| Number of pressure damages on duodenum | Equal variances assumed | 63.649 | .000 | -2.110 | 11 | .059 | -3.40476 | 1.61374 | -6.95658 | .14705 |
| | Equal variances not assumed | | | -2.290 | 6.152 | .061 | -3.40476 | 1.48709 | -7.02181 | .21229 |

## 8.5.11 Appendix 5.5.1 – Example scoring tool

**Station 8: Laparoscopic cholecystectomy**

| Trainee number | Trainer | Date |
|---|---|---|
| | | |

**Instructions to trainee:**

A 50-year-old female patient is undergoing a laparoscopic cholecystectomy for biliary colic. Please proceed as appropriate.

**Procedure-Based Assessment (Intra-Operative Domain)**

| Item | Description | Done | Not done |
|---|---|---|---|
| 1 | Follows an agreed, logical sequence or protocol for the procedure | | |
| 2 | Consistently handles tissue well with minimal damage | | |
| 3 | Controls bleeding promptly by an appropriate method | | |
| 4 | Demonstrates a sound technique of knots and sutures/ staples | | |
| 5 | Uses instruments appropriately and safely | | |
| 6 | Proceeds at appropriate pace with economy of movement | | |
| 7 | Anticipates and responds appropriately to variation (e.g.) anatomy | | |
| 8 | Deals calmly and effectively with unexpected events/ complications | | |
| 9 | Uses assistant(s) to the best advantage at all times | | |
| 10 | Dissects cholecystectomy triangle safely | | |
| 11 | Safely ligates and divides the cystic duct and cystic artery | | |
| 12 | Carefully mobilises gallbladder off the liver | | |
| 13 | Safely extracts gallbladder from a port site | | |

**NEXT PAGE**

**OSATS**

| **Respect for Tissue** | | Tick |
|---|---|---|
| 1. | Frequently and unnecessary force on tissue or caused damage by inappropriate use of instruments | |
| 2. | | |
| 3. | Careful handling of tissue but occasionally caused inadvertent damage | |
| 4. | | |
| 5. | Consistently handled tissue appropriately with minimal damage | |
| **Time and Motion** | | |
| 1. | Many unnecessary moves | |
| 2. | | |
| 3. | Efficient time/ motion but some unnecessary moves | |
| 4. | | |
| 5. | Clear economy of movement and maximum efficiency | |
| **Instrument Handling** | | |
| 1. | Repeatedly makes tentative or awkward moves with instruments by inappropriate use of instruments | |
| 2. | | |
| 3. | Competent use of instruments but occasionally appeared stiff or awkward | |
| 4. | | |
| 5. | Fluid moves with instruments and no awkwardness | |
| **Knowledge of instruments** | | |
| 1. | Frequently asked for the wring instrument of used inappropriate instrument | |
| 2. | | |
| 3. | Knew names of most instruments and used appropriate instrument | |
| 4. | | |
| 5. | Obviously familiar with the instruments and their names | |
| **Flow of operation** | | |
| 1. | Frequently stopped operating and seemed unsure of next move | |
| 2. | | |
| 3. | Demonstrated some forward planning with reasonable progression of procedure | |
| 4. | | |
| 5. | Obviously planned course of operation with effortless flow from one move to the next | |
| **Use of assistants** | | |
| 1. | Consistently placed assistants poorly or failed to use assistants | |
| 2. | | |
| 3. | Appropriate use of assistants most of the time | |
| 4. | | |
| 5. | Strategically used assistants to the best advantage at all times | |
| **Knowledge of specific procedure** | | |
| 1. | Deficient knowledge. Needed specific instruction at most steps | |
| 2. | | |
| 3. | Knew all important steps of operation | |
| 4. | | |
| 5. | Demonstrated familiarity with all aspects of the operation | |

**NEXT PAGE**

**Procedure-Based Assessment (Global Summary)**

| Level at which completed elements of the PBA were performed on this occasion | | Tick as appropriate |
|---|---|---|
| Level 0 | Insufficient evidence observed to support a summary judgment | |
| Level 1a | Able to assist with guidance (was not familiar with all steps of the procedure) | |
| Level 1b | Able to assist without guidance (Knew all steps of procedure and anticipated next move) | |
| Level 2a | Guidance required for most/ all of the procedure (or part performed) | |
| Level 3a | Procedure performed with minimal guidance or intervention (needed occasional help) | |
| Level 3b | Procedure performed competently without guidance of intervention but lacked fluency | |
| Level 4a | Procedure performed competently without guidance or intervention | |
| Level 4b | As 4a and was able to anticipate, avoid and/ or deal with common problems/ complications | |

**Generic Global Summary**

| Not Competent | Borderline | Competent |
|---|---|---|
| | | |

## 9. Corresponding Author Declarations

RCSI UNIVERSITY OF MEDICINE AND HEALTH SCIENCES

### Corresponding Author Declaration

This declaration describes the independent research contribution of the candidate and each corresponding author for each paper constituting the thesis.

For each paper the declaration should be completed and signed by the candidate and the corresponding author(s).

| Paper no. | Title |
|---|---|
| 1 | Assessing Operative Skill in the Competency-based Education Era: Lessons From the UK and Ireland |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Study conceptualisation. Search strategy/ methodology. Screening and full-text review. Data extraction. Data synthesis (narrative review). Writing – original draft, corrections post-internal and external review. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 2 | How do Trainers and Trainees use Formative Workplace-based Assessments of Operative Competence to Deliver and Receive Structured Feedback? |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Study conceptualisation. Methodology development. Statistical analysis and data curation. Writing – original draft, corrections post-internal and external review. Data visualisation. Project administration. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 3 | Defining operative experience targets in surgical training: A systematic review |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Study conceptualisation. Search strategy/ methodology. Screening and full-text review. Data extraction. Data synthesis tabulation. Writing – original draft, corrections post-internal and external review. Co-author supervision and project administration. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 4 | Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Study conceptualisation. Search strategy/ methodology. Screening and full-text review. Data extraction. Data synthesis (tabulation). Writing – original draft, corrections post-internal and external review. Co-author supervision and project administration. Data visualisation. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 5 | Perspectives on simulation-based assessment of operative skill in surgical training |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Methodology. Interview tool development. Conducting interviews. Transcription. Thematic analysis (data synthesis). Writing – original draft, corrections post-internal and external review. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 6 | Baseline Assessments of Psychomotor Aptitude are Associated with the Future Operative Performance of Surgical Trainees A Prospective Observational Cohort Study |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Study conceptualisation. Methodology development. Statistical analysis and data curation. Writing – original draft, corrections post-internal and external review. Data visualisation. Project administration. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 7 | Training and assessment using the LapSim laparoscopic simulator: a scoping review of validity evidence |

| Candidate: | |
|---|---|
| The independent contribution of the candidate. | Study conceptualisation. Search strategy/ methodology. Screening and full-text review. Data extraction. Data synthesis (tabulation). Writing – original draft, corrections post-internal and external review. Co-author supervision and project administration. Data visualisation. |

Signature of candidate

Signature of corresponding author

**Manuscripts in preparation/ under review**

| Paper no. | Title |
|---|---|
| 8 | The current status of simulation as a high stakes assessment modality in surgery |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Study conceptualisation. Search strategy/ methodology. Screening and full-text review. Data extraction. Data synthesis. Writing – original draft, corrections post-internal review. Co-author supervision and project administration. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 9 | Perceptions and experiences of simulation as an assessment modality in surgery – a survey of trainees |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Methodology. Survey item development. Data collation and analysis. Writing – original draft. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 10 | Perceptions and experiences of simulation as an assessment modality in surgery – a survey of trainees |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Methodology. Survey item development. Data collation and analysis. Writing – original draft. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 11 | The predictive validity of a national selection process in surgical training |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Study conceptualisation. Methodology development. Statistical analysis and data curation. Writing – original draft, corrections post-internal review. Data visualisation. Project administration. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 12 | . Exploring the number of observations required to reliably assess trainees in Core Surgical Training |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Study conceptualisation. Methodology development. Statistical analysis and data curation. Writing – original draft, corrections post-internal review. Data visualisation. Project administration. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 13 | Comparing the reliability of workplace-based and simulation-based assessments of operative competence |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Study conceptualisation. Methodology development. Statistical analysis and data curation. Writing – original draft, corrections post-internal review. Data visualisation. Project administration. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 14 | Generating a prioritised list of operative procedures for simulation-based assessment of General Surgery Trainees through consensus |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Study conceptualisation. Methodology development. Statistical analysis and data curation. Writing – original draft, corrections post-internal review. Data visualisation. Project administration. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 15 | Development and validation of a novel simulation-based assessment of operative competence for higher specialist trainees in General Surgery |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Study conceptualisation. Methodology development. Data collection. Recruitment. Model development. Statistical analysis and data curation. Writing – original draft, corrections post-internal review. Data visualisation. Project administration. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 17 | Examining the acceptability of a novel simulation-based assessment of operative competence in Higher Specialist Training – A qualitative analysis |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Methodology. Interview tool development. Conducting interviews. Transcription. Thematic analysis (data synthesis). Writing – original draft, corrections post-internal and external review. |

Signature of candidate

Signature of corresponding author

| Paper no. | Title |
|---|---|
| 18 | Simulation-based assessment across the continuum of technical skill acquisition in surgical training from selection to certification: an integrative framework |

| Candidate: | |
|---|---|
| The independent contribution of the candidate | Writing – original draft, corrections post-internal review. |

Signature of candidate

Signature of corresponding author