Lund University

Surgical Education, Evaluating selection methods and early training experiences in surgical residency

Pedersen, Hanne

2025

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Pedersen, H. (2025). *Surgical Education, Evaluating selection methods and early training experiences in surgical residency*. [Doctoral Thesis (compilation), Department of Clinical Sciences, Lund]. Lund University, Faculty of Medicine.

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Download date: 29. Jun. 2026

Surgical Education

Evaluating selection methods and early training experiences in surgical residency

HANNE PEDERSEN
DEPARTMENT OF CLINICAL SCIENCES, LUND | FACULTY OF MEDICINE | LUND UNIVERSITY

Surgical Education

Evaluating selection methods and early training experiences in surgical residency

Hanne Pedersen

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on 21th of November at 13.30 in Belfragesalen, BMC, Lund

Faculty opponent
Professor Li Felländer-Tsai
Karolinska Institutet, Stockholm, Sweden

Organization: LUND UNIVERSITY

Document name: Doctoral Dissertation

Date of issue: 2025-11-21

Author: Hanne Pedersen

Sponsoring organizations:
Prostatacancerförbundet
Löf (Löf regionernas ömsesidiga försäkringsbolag)
Vera och Carl J. Michaelsens Donationsfond
Region Skåne

Title and subtitle: Surgical Education, Evaluating selection methods and early training experiences in surgical residency

Abstract

Background Selection of surgical residents is carried out in a variety of ways across the globe. In Sweden, the selection process is characterized by a lack of transparency and is conducted locally at each hospital. Recruiting suitable candidates is a matter of patient safety since patient outcome is related to surgeon characteristics and behaviors.

Aim The aim of this thesis is to explore quantitative and qualitative assessment instruments for technical and non-technical skills in surgical residents at the time of recruitment/early in the career.

Methods The first paper utilized Google and PubMed to identify selection processes in the European Union (EU). Board delegates in the Section of Surgery of the Union Européenne des Médecins Specialistes (UEMS), were contacted requesting pertinent information. The following three studies were conducted using the same cohort of 50 applicants for a residency or locum position in general surgery, urology, or pediatric surgery from seven hospitals. Participants were inquired for background factors, and underwent evaluations including visuospatial ability, laparoscopic performance, personality ratings, 360-degree evaluations, and interviews.

Results A multitude of recruitment strategies were identified across the EU. Visuospatial ability correlated to performance in a laparoscopic simulator. Extravert participants received higher ratings on the 360-degree evaluations, including technical skill, all though no correlation to objective measures of technical skill was identified. The interviews revealed a cautious feedback culture with a reluctancy to provide feedback.

Significance The first study demonstrated different traditions regarding the best recruitment process, including whether the resident is seen as part of higher education or as an employee. The test of visuospatial ability might serve as a promising indicator of technical performance. The use of 360-degree evaluations should be used with caution to make decisions/summative feedback concerning selection. Insights into the current feedback culture may serve as a basis for targeted improvements. The findings may be relevant to an international audience in postgraduate medical education.

Key words: Surgical education, Selection, Technical skills, Non-technical skills, Feedback, Personlity, Visuospatial ability, Learning environment, Multisource feedback

Language: English

Number of pages: 73

ISSN and key title: 1652-8220, Lund University, Faculty of Medicine Doctoral Dissertation Series 2025:126

ISBN: 978-91-8021-779-8

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature

Date 2025-10-06

Surgical Education

Evaluating selection methods and early training experiences in surgical residency

Hanne Pedersen

Coverpicture by Isabelle Pedersen

Copyright pp 1-73 Hanne Pedersen

Paper 1 © BMC Medical Education

Paper 2 © Surgery Open Science

Paper 3 © Journal of Surgical Research

Paper 4 © by the Authors (Manuscript submitted)

Faculty of Medicine

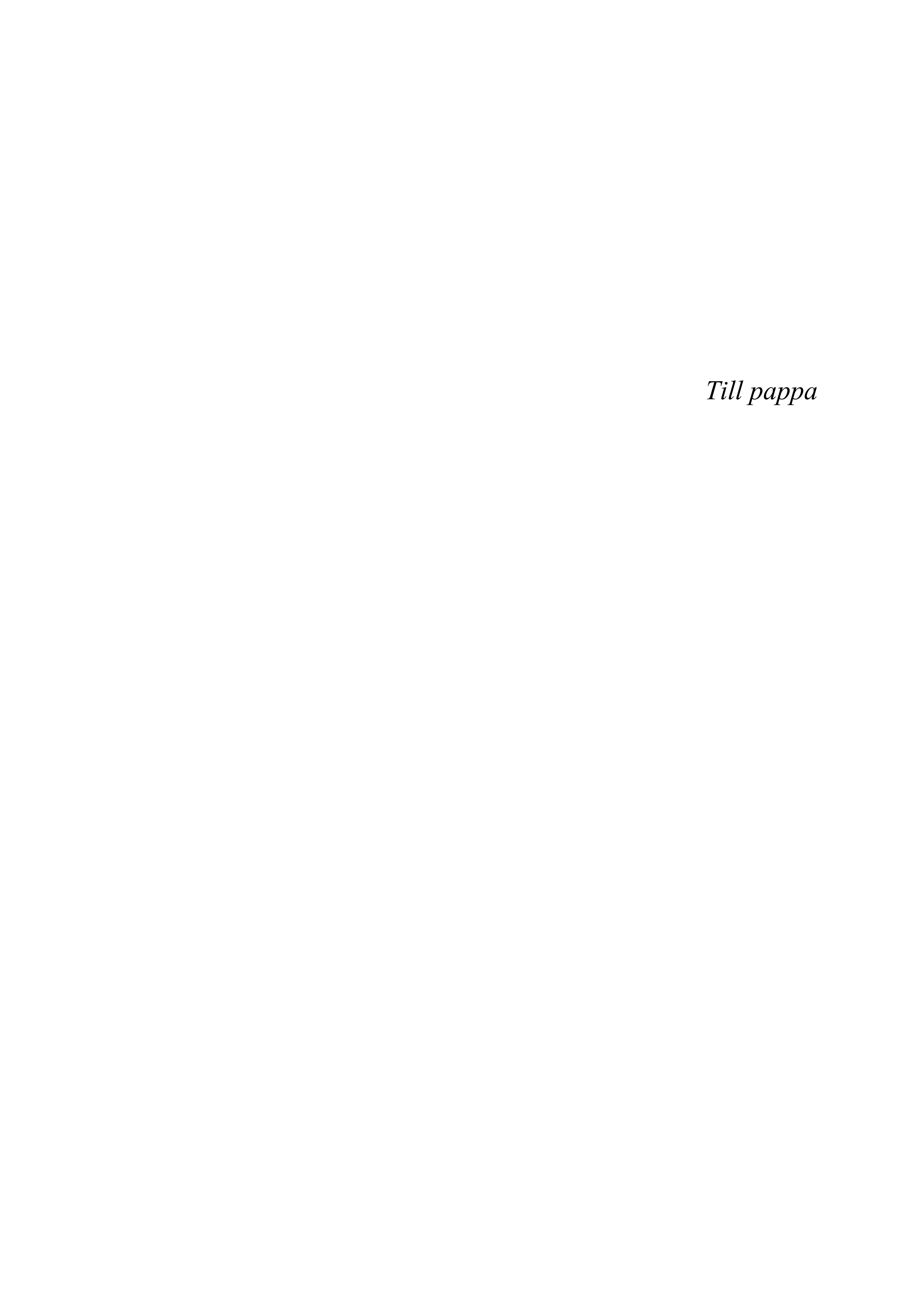
Department of Clinical Sciences, Pediatrics, Lund, Sweden

ISBN 978-91-8021-779-8

ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University

Lund 2025



Till pappa

Table of Contents

Thesis at a glance 9
Abstract 10
Populärvetenskaplig sammanfattning 11
List of papers 15
Abbreviations 16

Introduction 17
The surgeon and patient outcomes 17
Attrition and remediation during residency 18
Selection processes 19
Technical skill 21
Non-technical skill 22
Personality 22
Unsuitability 24
Feedback 25
Summary and aims 26

Methods and material 27
Study design and participants 27
Paper I 28
Paper II 29
Paper III 32
Paper IV 33
Qualitative and quantitative analysis 34
Ethical considerations 35

Results 36
Paper I 36
Participants Paper II-IV 37
Paper II 39
Paper III 40
Paper IV 42

Discussion **46**
Recruitment process and prediction of job performance............................46
Technical skill 48
Non-technical skill and halo effect..49
Assessment, feedback and the learning environment.................................51
Conclusion **53**
Future perspectives...54
Acknowledgements **59**
Errata **61**
References **62**

Thesis at a glance

	Aim	Method	Results	Conclusion
I	Investigate and compare selection processes to surgical residency training in the European Union (EU)	Information on selection methods in the EU was obtained. Comparative case study between a country with a decentralized process and a country with a centralized process.	A multitude of recruitment strategies were identified. Countries were divided on a scale from a decentralized to a centralized process.	There is a wide variation in selection processes in the EU. Centralized and decentralized processes hold different advantages.
II	Investigate the impact of demographic factors, manual dexterity, and visuospatial ability on the performance in two different laparoscopic simulators amongst applicants to surgical resident positions	Fifty applicants for a residency or locum position in a surgical specialty were included. Background factors, visuospatial ability (VSA), dexterity, and laparoscopic performance were investigated. Logistic and linear regression.	A significant association was found between VSA and a 2D simulator. With every point on the VSA test the odds of accomplishing the task increased by 11 %.	VSA could potentially be used as a selection method in the future and can be used to individualize laparoscopic training programs.
III	Investigate the relationship between personality and ratings on 360-degree evaluations among surgical residency applicants	The same cohort as in Paper II. Results from laparoscopic simulators, personality ratings, and 360-degree evaluations were included. Pearson correlations and Student's *t*-test.	The participant rated higher in four of five personality domains compared to norm values. Extravert persons were rated higher on 360-degree evaluations, without correlations to objective measures of technical skills.	The personality among aspiring surgeons differs from the norm. Objective measures for assessment should be used, and assessors should be trained in assessment to avoid bias.
IV	Explore experiences and challenges in handling feedback and criticism among early career surgeons seeking surgical residencies	The same cohort as in Paper II. Participants were interviewed using a semi-structured interview-guide. A cross-sectional thematic analysis was performed.	Four themes were constructed 1) reflection and processing of criticism, 2) emotional response to criticism, 3) cautious feedback culture, and 4) navigating criticism in a hierarchical system.	A gap in the feedback culture was described with a reluctancy to give feedback. Suggestions for improvement are made.

// Abstract

Background

Selection of surgical residents is carried out in a variety of ways across the globe. In Sweden, the selection process is characterized by a lack of transparency and is conducted locally at each hospital. Recruiting suitable candidates is a matter of patient safety since patient outcome is related to surgeon characteristics and behaviors.

Aim

The aim of this thesis is to explore quantitative and qualitative assessment instruments for technical and non-technical skills in surgical residents at the time of recruitment/early in the career.

Methods

The first paper utilized Google and PubMed to identify selection processes in the European Union (EU). Board delegates in the Section of Surgery of the Union Européenne des Médecins Specialistes (UEMS) were contacted requesting pertinent information. The following three studies were conducted using the same cohort of 50 applicants for a residency or locum position in general surgery, urology, or pediatric surgery from seven hospitals. Participants were inquired for background factors, and underwent evaluations including visuospatial ability, laparoscopic performance, personality ratings, 360-degree evaluations, and interviews.

Results

A multitude of recruitment strategies were identified across the EU. Visuospatial ability correlated to performance in a laparoscopic simulator. Extravert participants received higher ratings on the 360-degree evaluations, including technical skill, all though no correlation to objective measures of technical skill was identified. The interviews revealed a cautious feedback culture with a reluctancy to provide feedback.

Significance

The first study demonstrated different traditions regarding the best recruitment process, including whether the resident is seen as part of higher education or as an employee. The test of visuospatial ability might serve as a promising indicator of technical performance. The use of 360-degree evaluations should be used with caution to make decisions concerning selection. Insights into the current feedback culture may serve as a basis for targeted improvements. The findings may be relevant to an international audience in postgraduate medical education.

Populärvetenskaplig sammanfattning

Bakgrund

En kirurgs karriär startar vid rekryteringen. Efter avslutad allmäntjänstgöring (AT) kan den legitimerade läkaren söka specialiseringstjänstgöring (ST), inom det område som önskas, exempelvis inom kirurgi. ST är en utbildningstjänst, där doktorn utvecklas och tränas in i sin valda specialitet. En ST-tjänst varar minst fem år och avslutas med en godkänd ansökan om specialistbevis.

Att välja en lämplig kandidat redan vid anställning är en fråga om patientsäkerhet eftersom utfallet vid en operation bland annat beror på vem som opererar. Faktorer så som kön, teknisk förmåga, kirurgisk volym och kommunikation har visat sig påverka utfallet för en patient. Detta betyder att både teknisk och icke-teknisk förmåga är viktigt för att bli en kompetent kirurg. Icke-tekniska förmågor är förmågor som behövs vid komplexa uppgifter, till exempel beslutsfattande om man ska operera eller ej, val av behandling och vid kommunikation med såväl medarbetare som patienter och anhöriga. Träning avseende icke-tekniska förmågor har på senare tid blivit allt mer implementerad inom sjukvården, till exempel med interprofessionell teamträning.

I Sverige drabbades 2023 drygt var tionde (10,7 %) av alla patienter i sjukvården av en skada i samband med vård. Skador efter kirurgi eller annan invasiv åtgärd stod för 11,9 % av skadorna. Alla är inte undvikbara, och en vårdskada definieras enligt patientsäkerhetslagen som ”lidande, kroppslig eller psykisk skada eller sjukdom samt dödsfall som hade kunnat undvikas om adekvata åtgärder hade vidtagits vid patientens kontakt med hälso- och sjukvården”. Vårdskador leder förutom till lidande för patienter, och ibland även till död, också till ökade kostnader med förlängda vårdtider och i vissa fall till ytterligare operationer. År 2023 bedömdes 41, 1 % av de kirurgiska skadorna vara undvikbara.

Inom kirurgin har rekryteringen föreslagits vara ”den saknade länken” i patientsäkerhetsarbetet. Den nuvarande rekryteringen världen över sker på en mängd olika sätt och med olika förutsättningar. Vissa länder har ett strukturerat och transparant urval, och andra länder har lokala processer. I Sverige sker rekryteringen lokalt, saknar transparens och det finns inga riktlinjer som stöd. En vanlig anställningsform är ett vikariat på 6–12 månader med möjlighet till utvärdering av den sökande under en längre period innan anställningen förlängs till en fast ST-tjänst. Inom andra organisationer som arbetar med höga risker för liv, som inom flygindustrin, är rekryteringen standardiserad sedan länge. Inom flygindustrin har attityder med potentiell inverkan på omdöme och beslutsfattande identifierats, och piloter genomgår utbildning för att utveckla medvetenhet och sin förmåga att korrigera dessa attityder. I en studie på ortopediska kirurger har en av dessa attityder visat sig korrelera till ökad risk för re-operation.

Flera försök har gjorts för att identifiera faktorer som kan förutse hög prestation hos kirurger. Faktorer så som kön, visuospatial förmåga, fingerfärdighet och datorspelande har förslagits, men varken någon särskild faktor eller metod har visat sig vara tillräckligt prediktiv. Studier har visat skillnader mellan kirurger och icke-kirurger avseende personlighet, men med variationer mellan olika studier och länder. Ett vanligt mått på en lyckad rekrytering är avhopp, och i andra länder har detta beskrivits som en stor utmaning med varierande antal, upp till så högt som 40 %. Siffror gällande hur många som fullföljer en ST-utbildning i Sverige saknas. Däremot uppgav 85 % av erfarna kirurger i en svensk studie att de träffat på en ST-läkare de bedömt vara olämplig.

Syfte

Den här avhandlingen är en del av en longitudinell studie som följer och utvärderar ST-läkare inom kirurgiska discipliner från rekrytering tills att de får sitt specialistbevis. Målet med den här avhandlingen är att utvärdera kvantitativa och kvalitativa metoder som mäter både teknisk och icke-teknisk förmåga hos sökande till en ST-tjänst inom en kirurgisk disciplin.

Metod

Totalt fyra studier är inkluderade i avhandlingen. I Studie I undersöktes hur rekryteringen såg ut i länder i Europeiska unionen (EU). Google och PubMed användes för informationssökande. Email skickades till styrelsemedlemmar i den kirurgiska delen av Union Européenne des Médecins Specialistes (UEMS) med frågor kring deras respektive rekryteringsprocess. Djupare information kring Irlands och Sveriges rekrytering samlades in genom intervjuaanalys och närvaro vid Royal College of Surgeons’ rekryteringsdagar på Irlands.

Studie II- IV inkluderade samma kohort av 50 sökande till en ST-tjänst eller ett vikariat i allmänkirurgi, urologi eller barnkirurgi från sju sjukhus. Inklusionen av deltagare skedde mellan 2017–2021 genom att deltagande kliniker skickade kontaktinformation till forskargruppen i samband med att de intervjuade personer för en tjänst. Deltagarna bokades in för en heldag av tester. Samtliga ombads fylla i en bakgrundsenkät med information kring exempelvis kön och ålder. Teknisk förmåga testades genom övningar i två olika laparoskopiska (titthålskirugiska) simulatorer, samt genom test av visuospatial förmåga och fingerfärdighet. Icke-teknisk förmåga utvärderades genom personlighetsskattning, 360-graders bedömningar och en semi-strukturerad intervju. Studie II-III använde kvantitativa analyser (siffror och statistik) och Studie IV använde kvalitativ metod med tematisk analys av intervjuer.

Resultat

I Studie I identifierades rekryteringsprocessen i 27 länder. Dessa delades in på en skala från centraliserat till decentraliserat urval. Irland hade den mest utvecklade

rekryteringsprocessen och var det enda landet som testade teknisk förmåga. Sverige valdes ut som ett exempel från den decentraliserade sidan av skalan och en jämförelse länderna emellan utfördes.

I Studie II analyserades data från de två laparoskopisimulatorerna, visuospatial förmåga, bakgrundsenkäten och fingerfärdighetstestet. Det fanns en signifikant relation mellan visuospatial förmåga och resultatet i en tvådimensionell laparoskopisimulator. Oddsen för att klara av simulatoruppgiften ökade med 11 % för varje poäng på testet av visuospatial förmåga. Det fanns inga andra signifikanta korrelationer mellan utfallet i simulatorerna och någon av bakgrundsfaktorerna eller fingerfärdighetstestet.

I Studie III analyserades personlighetsskattningarna, 360-graders bedömningar och resultatet från de laparoskopiska simulatorerna. Jämfört med normalvärden för personlighet skattade den inkluderade kohorten högre på fyra av fem personlighetsdomäner utifrån den vanligt använda Femfaktorteorin: öppenhet, emotionell stabilitet, samvetsgrannhet och vänlighet. Det var enbart på extraversion deltagarna inte skattade sig högre. 360-graders bedömningarna var generellt höga, och totalt sex korrelationer fanns till personlighet. De två starkaste korrelationerna var till extraversion. Det fanns inga signifikanta korrelationer mellan personlighet och resultaten i de laparoskopiska simulatorerna.

Studie IV var en kvalitativ studie som utvärderade erfarenheter och utmaning med att hantera feedback och kritik i Sverige. Deltagarna beskrev en brist på feedback, framför allt de deltagare som hade arbetat i ett annat land och kunde jämföra mellan länderna. Flera av deltagarna upplevde också att det var svårt att ge feedback och kritik, men några ställde sig positiva till att lära ut. Strategierna för att hantera kritik skiljde sig åt, men ett vanligt sätt att hantera kritik bland deltagarna var att analysera och reflektera över den själv innan de valde att acceptera den. Flera deltagare beskrev även en arbetsmiljö där orättvis kritik var förekommande. En del upplevde att det var svårt att ta emot kritik och att det kunde påverka deras mående. Den hierarkiska miljön påverkade också möjligheten att ge kritik till någon senior, av rädsla för att bli straffad.

Slutsatser

Det finns stora variationer i hur rekryteringen av ST-läkare i kirurgi ser ut och inom EU hade inga länder en exakt likadan metod. Förutsättningarna skiljer sig åt och en modell fungerar inte för alla på grund av till exempel olika anställningsmodeller och ekonomiska förutsättningar. I Sverige finns förbättringspotential för att arbeta mer evidensbaserat vid rekrytering genom att använda sig av en strukturerad intervjumall, vara flera som intervjuar tillsammans och genom att intervjua flera sökanden. Avseende teknisk förmåga är test av visuospatial förmåga mest lovande som surrogatmått och skulle potentiellt kunna bli en del av en framtida

rekryteringsprocess. Det skulle även kunna användas för att anpassa inlärning av titthålskirurgi till en individuell nivå.

Den kirurgiska personligheten är svårdefinierad men skiljer sig från normalpopulationen då deltagarna skattade sig högre på fyra av fem personlighetsdomäner. På 360-graders bedömningar från medarbetare korrelerade höga bedömningar med extraversion, vilket pekar på en potentiell bias och vikten av att träna bedömarna för att få korrekta värderingar. Det finns förbättringspotential i den svenska feedbackkulturen som unga doktorer befinner sig i då unga läkare inte uppfattar sig få tillräckligt med konkret återkoppling på saker de kan förbättra. Det finns dessutom en tveksamhet i att ge feedback även från unga doktorers sida. Förslag till förbättringar i lärandemiljön inkluderar till exempel att schemalägga nya doktorer i sammanhängande perioder med sina handledare under ST-tjänstgöringen, att redan vid början av anställningen tydliggöra vilka förväntningar som föreligger rörande kompetensmål att uppnå inom de första 6 – 12 månaderna, samt att alla kollegor utbildas i pedagogik för handledning och feedback.

Slutligen har denna avhandling initierat ett arbete med det långsiktiga målet att göra rekryteringen av ST-läkare inom kirurgi mer evidensbaserat. Genom förbättrad rekrytering och utbildning skulle patientsäkerheten kunna ökas.

List of papers

I. Different approaches to selection of surgical trainees in the European Union

   Kristine Hagelsteen, **Hanne Pedersen**, Chris Mathieu

   BMC Medical Education 2021

II. Visuospatial ability is associated to 2D laparoscopic simulator performance amongst surgical residents

   Hanne Pedersen, Darya Ståhl, Mikael Ekelund, Magnus Anderberg, Martin Bäckström, Anders Bergenfelz, Kristine Hagelsteen

   Surgery Open Science 2023

III. Extravert Surgical Resident Applicants Get Higher 360-Degree Evaluations From Coworkers

   Hanne Pedersen, Martin Bäckström, Kristine Hagelsteen

   Journal of Surgical Research 2024

IV. Feedback gap and strategies for handling criticism in early surgical career

   Hanne Pedersen, Alexander Tejera, Christopher Mathieu, Britt-Marie Johansson, Magnus Anderberg, Kristine Hagelsteen.

   Manuscript.

Abbreviations

2D	Two dimensional
3D	Three dimensional
VSA	Visuospatial ability
VR	Virtual Reality
NASCE	Network of Accredited Clinical Skills Centres in Europe
NB5I	Neutralized Big Five Inventory
REDCap	Research Electronic Data Capture tools hosted by Lund University.
UEMS	Union Européenne des Médecins Specialistes
RCSI	Royal College of Surgeons in Ireland
MMI	Multiple Mini Interviews
CRM	Crew Resource Management
BT	Bastjänstgöring
ST	Specialiseringstjänstgöring
AI	Artificial intelligence
HRO	High Reliability Organizations
CST	Core Surgical Training
HST	Higher Specialist Training
SJT	Situational Judgement Tests
SYLF	Sveriges Yngre Läkares Förening

Introduction

The career of a surgeon commences at the point of recruitment. After completing an employment as an intern, the doctor applies for a residency position, a training position in their preferred specialty to become a certified specialist. During the period of surgical training, surgeons develop and refine their skills, and if successful, become specialists in their respective fields. The selection of suitable candidates is a matter of patient safety, given that patient outcomes vary depending on the surgeon. Both technical and non-technical skills, as well as other factors, influence the outcome [1,2]. In Sweden, surgery accounted for 11.9 % of adverse events in 2023. Despite a recent decline, still 41.1 % of surgical adverse events were deemed avoidable [3]. Moreover, a study from the United States have documented that the attrition rate among surgical residents varies, with rates that can reach up to 40 %. Attrition has significant financial implications, where the cost of medical education for a resident was found to be 80,000 dollars per year. This cost did not include costs such as onboarding or costs related to interviewing [4]. The present Swedish selection process is conducted locally at each hospital depending on local staffing needs, it lacks transparency, and does not include guidelines for selecting the most suitable candidate [5]. Limited research has been conducted on surgical residency in the Swedish context. By optimizing recruitment, patient safety could potentially be enhanced, and the cost could be reduced, allowing for greater investment in patient care.

The surgeon and patient outcomes

The relationship between a surgeon and a patient is special and built on trust. Complications are unwanted but common overall in healthcare, and not all are preventable. Adverse events are estimated to be affecting 10 % of hospitalized patients worldwide in high-income countries, and accounting for up to 15 % of hospital costs [6]. Adverse events result in patient suffering and sometimes death. In addition to patient suffering, there are additional monetary costs such as prolonged hospital stays and re-operations [7]. Patient outcomes are influenced by who performed the operation, and factors such as sex [8-10], technical skill [1,11], surgical volume [12,13], and team behavior and communication [2,14-16] have been associated with outcome [1,2,8-12]. In a large Swedish study of over 150,000 patients undergoing cholecystectomy, male surgeons were associated with more surgical complications,

overall complications, and longer hospital stays [9]. A systematic review and meta-analysis found lower mortality with female surgeons, but no differences in readmission or complication rates. Possible contributing factors have been suggested, including that female surgeons may be more risk averse, and better at following guidelines [8]. Technical skill has been found to affect the surgical outcome. A wide variation in technical skill has been shown among bariatric surgeons, and those in the bottom quartile had significantly higher complication rates and higher mortality [11]. Similar results were found regarding surgeons' technical skill and outcome after colectomy, where higher skills were associated with lower complication rates, unplanned reoperations, and a composite of death or serious morbidity [1]. Few have studied the subject, and a systematic review conducted in 2023 included seven articles, including a total of 151 surgeons. The subsequent meta-analysis showed significant associations between surgical skill and reoperations, obstruction, hemorrhage, and any medical complication [17]. Further, increased surgical volume among surgeons performing radical prostatectomies for localized prostate cancer were associated with better results, including lower mortality and lower perioperative complications [13]. Non-technical skills have also been found to have an impact on surgical outcome. Impaired team behavior increases the odds of complication or death [2], and surgeons with unprofessional behavior, defined as receiving a higher number of coworker reports on the subject, have been found more likely to suffer from any type of complication [14].

Attrition and remediation during residency

Every recruitment is a future investment for the employer, and with the expenses and resources following a position as a resident the employer strives for a high return of investment. A comprehensive selection process incurs considerable expenses [4,18], with additional costs arising from attrition or remediation [4]. Attrition is a pervasive challenge, with estimates ranging up to as high as 40 % [4]. However, studies in general surgery show great variation, and a systematic review and meta-analysis yielded a pooled estimate of 18 % [19]. General surgery has higher attrition rates than other surgical subspecialties such as urology [20]. Besides the monetary costs, attrition leads to fewer specialists and prolonged time to becoming a specialist for the individual doctor. The majority of residents leave during the initial two years of their training, with various factors contributing to this phenomenon. Frequently cited factors are lifestyle considerations, health reasons, familial or personal obligations, financial burden, and opportunities for earlier specialization [21]. Remediation is a prevalent occurrence, and a retrospective study on residents in general surgery revealed a rate of approximately 30 % requiring at least one remediation [22]. There is a lack of Swedish data regarding both attrition and remediations and most studies have been focused on Anglo-Saxon countries.

Selection processes

The validity of different assessment methods from 85 years of research was summarized in 1998 by Schmidt and Hunter. The validity of 19 selection procedures was investigated, and the included methods ranged from General Mental Ability to graphology. Three combinations of methods had the highest validity: 1) general mental ability and work sample test, 2) general mental ability and an integrity test, and 3) general mental ability and a structured interview [23]. A structured interview using an interview guide has higher reliability than an unstructured one [23,24].

In other High Reliability Organizations (HRO) where human lives are at stake, such as the aviation industry, the recruitment and employment of a standardized selection process has been a long-standing practice [25]. In the field of surgery, the selection of surgical residents has been posited to represent "the missing link" in patient safety [26]. The current selection of surgical residents worldwide is conducted in a multitude of ways, with some countries employing a structured and transparent selection process, and others utilizing local processes. In Ireland, the recruitment of surgeons consists of two separate selections. The first selection is for Core Surgical Training (CST), which lasts two years, and a second for Specialty Training, lasting six years. Not all will be able to proceed to surgical training at the second stage. For both selections a meticulous and systematic approach is employed to evaluate candidates with explicit guidelines and a standardized evaluation framework. For CST these evaluations include both technical skills, such as visuospatial ability, and non-technical skills, such as communication [27,28]. During covid-19 the method for CST underwent modifications into being completely conducted virtually. Conversely, in countries such as Sweden, the selection process is characterized by being local and often nontransparent, with a paucity of common guidelines available to assist in the selection process [5]. A common evaluation method is to be employed as a locum for 6-12 months, thereby allowing for continuous evaluations before progressing into a permanent position as a resident [29].

In Sweden, the recruitment of doctors received some increased interest in the last years, and in 2018 Sveriges Yngre Läkares Förening (SYLF) started "Schysst Rekrytering!", which translates into "Fair Recruitment!". SYLF is a union young doctors that belongs to The Swedish Medical Association, the trade union for medical doctors in Sweden. "Schysst Rekrytering!" was developed for the recruitment of interns, with specific criteria to be fulfilled by the employer (Table 1). The employer should use a structured and reproducible method, and the process should be designed in a way that ensures fairness and equitable treatments for all individuals, regardless of factors such as sex, age, origin, or other similar criteria. [30]. No similar guidelines are available for recruitment of residents.

Table 1. Sveriges Yngre Läkares Förening (SYLF) critiera for Schysst Rekrytering! (fair recruitment) among interns.

✓	The employer is able to show that the criteria outlined in the advertisement are being used to select candidates for interviews and job offers
✓	The employer utilizes a structured and reproducible method for conducting interviews and other tests
✓	SYLF has the possibility to nominate a representative that on paid working hours may elect to participate in interviews and tests as deemed necessary
✓	One representative of SYLF is permitted to participate in the selection process for interviews and the final selection during paid working hours

In the selection processes worldwide a broad array of evaluation methods are employed, encompassing for example academic performance [31], letters of recommendation [31,32], research output [31], and interviews [31,33]. The interview formats demonstrate significant variability [33], despite that structured interviews offer the most robust evidence [23,24,34]. Other potential interview formats that have been used include Multiple Mini Interviews (MMI), which also are based on structured questions, and Situational Judgment tests (SJT). MMIs consist of several short interviews with different assessors evaluating the interviewee at each station. At each station there is one interviewer or observer, and the station may include for example discussion on ethical decision making, or communications skills [35]. MMI has been shown reliable for assessment [36-38]. SJT, in contrast, are usually not performed in-person. It aims to assess the judgment and behavior of a candidate using specific scenarios from a given role. It has been used in some countries for both medicals school admission and for recruitment of residents [39,40].

Several attempts have been made in trying to predict future performance from information at recruitment, but with no current consensus on the best method [41-44]. During the study period for this thesis, challenges arose due to the covid-19 pandemic, which affected the conditions for selection and lead to a decrease in cost on an international level by the implementation of virtual interviews [38,45,46]. Instead, the pandemic gave rise to other potential challenges, including technical aspects [45]. While the extant research is limited, a review showed high satisfaction with virtual interviews [46]. In light of these findings, some have suggested that virtual interviews could become the new norm [47].

Technical skill

Technical skill is important for surgeons and studies have shown a correlation between technical skill and patient outcomes in terms of complications, reoperations and death [1,11]. Testing of technical skills has been suggested as part of resident selection [43,48], but is not commonly used today [49]. With the introduction of laparoscopic surgery, this technique must be learned, and studies suggest differences in learning curves and that some individuals may not be able to reach proficiency levels [48,50-53]. Laparoscopic simulators are used for training, and the skills acquired appear to be transferable to the operating room [54-57]. Both virtual reality simulators and box trainers appear to be similarly effective for training [58]. A difference has been shown between 2D and 3D simulators, with 3D simulators demonstrating improved laparoscopic performance [59-62] . Several attempts have been made to identify factors predictive of surgical aptitude, but there is no consensus on the best assessment method [42,43,49,63,64]. Factors that have been suggested to influence laparoscopic performance include for example sex, visuospatial ability, psychomotor skills, playing video games, or playing a musical instrument [43,63,65]. No single test has been shown to be reliable as a surrogate measure [43]. In a systematic review by Mason et al, the most promising factors include psychomotor skills, visuospatial ability (VSA), video game experience, and perceptual ability [65].

Some of the potential surrogate measures of laparoscopic performance are VSA or manual dexterity [66]. Visuospatial ability is defined as "*the skill of mentally manipulating an object in three dimensions*" [67], and it has been suggested to be fixed in adulthood [68]. This would make it an interesting proxy for technical ability, but evidence is limited, and some studies have shown improvements with training [69,70]. Two scoping reviews on VSA have been conducted recently. Kalun et al. found a positive association between simulated surgical performance and VSA [71]. In the other review, Maurice-Ventouris et al. conclude that a variety of different tests were used in the included studies, making comparisons and interpretation difficult [70].

Some studies have shown promising results between manual dexterity and laparoscopic performance [66,72,73]. Manual dexterity tests have been used for years to evaluate healthcare professionals, including surgeons. Several different tests have been used, with some of the most common being the Purdue Pegboard, the Grooved Pegboard Test, and the O´Connor Tweezer Dexterity Test [73]. Manual dexterity has been suggested to influence laparoscopic performance and has been shown to correlate with skill acquisition [64,72]. Additionally, the Purdue Pegboard has been correlated with initial endovascular skill among medical students [74]. Other background factors that have been suggested to influence laparoscopic performance include sex, gaming experience, previous laparoscopic experience, and playing music with somewhat conflicting results [63,65,75,76].

Non-technical skill

Non-technical skills (NTS) are skills that are needed in addition to the technical skills to fulfil the complex tasks of a surgeon, both in- and outside the operating theatre. Non-technical skills can be defined as "*a constellation of cognitive and social skills, exhibited by individuals and teams, needed to reduce errors and improve human performance in complex systems"*. NTS include factors such as situation awareness, decision making, teamwork and leadership, and communication. The training of non-technical skills originated from Crew Resource Management (CRM) in aviation which was developed in the 1980s. NTS, or the lack thereof, were identified as contributing factors in aviation accidents and subsequently integrated into pilot training [77]. Five Hazardous Attitudes have been identified in aviation: anti-authority, impulsivity, invulnerability, macho, and resignation. These attitudes are thought to impair judgement and decision making, and pilots are trained to both be aware of, and correct these behaviors [25]. Surgery is similar to aviation in many ways, with high demands on decision making and stress tolerance. In a study by Bruinsma et al. a modified questionnaire that in its original form was applied within aviation to identify hazardous attitudes among pilots was used. The modifications included for example an exchange of the word "pilot" to "surgeon". In their study on orthopedic surgeons 38 % had dangerously high levels of at least one of the hazardous attitudes [78]. In another study by Kadzielski et al. on orthopedic surgeons, one of the measured attitudes, the level of macho attitude, predicted almost 20% of the orthopedic surgeons' reoperation and readmission rates [79].

In healthcare, the importance of non-technical skills and patient safety was recognized in the 1990s and has been of increasing interest since then. Studies have shown that deficiencies in non-technical skills, such as communication failure, contribute to adverse events [77]. Training of non-technical skills, like communication, is to an increasing extent integrated into healthcare, for example with interprofessional team training education during all levels of the continuum of medical education [14,77,80]. Based on a prospective cohort study investigating non-technical skills in senior residents, demonstrated that senior residents were rated higher than junior residents and interns, which may demonstrate a development in non-technical skills during a surgical career [81].

Personality

Personality ratings are commonly used in recruitment in other professions and have been suggested to be a part of residency selection [82,83]. Personality relates to how individuals think, feel, and act. Personality traits are thought to be stable over time [84], although this still remains an ongoing debate [85]. There are several inventories used for assessment. The most commonly used personality theory is the Five-Factor

Model, also known as the Big Five. The Big Five was developed and acknowledged in the 1980´s and includes the domains of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (or Emotional Stability) (Figure 1). Each domain has its own specific characteristics. Openness is usually defined as being intelligent, curious, and imaginative. In contrast, those with low ratings would favor conservative values. High conscientiousness implies being organized and thorough, and a low score implies being more impulsive. Those rating high in extraversion are considered talkative, outgoing, and warm. On the contrary, those with low ratings in extraversion would be reserved and shy. Agreeableness concerns humanity, with those rating high being nurturing and caring, and those on the other spectrum being self-centered and indifferent to others. Neuroticism concerns experiences of distress, where higher ratings are associated with nervous tension, low self-esteem, and ineffective coping. Individuals with lower ratings are instead calm and even-tempered [86].

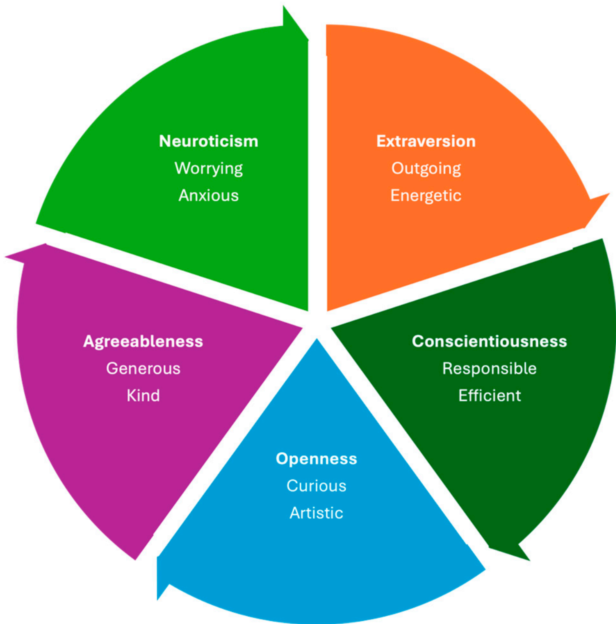


Figure 1. The domains of the Big Five including examples of traits under each domain.

Differences in personality have been shown between physicians of different medical specialties [87-89], and have been suggested to be related to choice of specialty [89]. A specific surgical personality profile has been discussed for years, and differences have been described both in comparison to other medical specialties and to the normative population. Despite this, the results have been somewhat mixed, and to date there is no consensus on a specific surgical personality profile [89-93]. A cross-cultural study between the United Kingdom and the Netherlands of gastrointestinal

surgeons demonstrated differences, such as for example the Dutch surgeons rating higher in openness. The variation in personality profile has been suggested to be influenced by culture and differences between surgical specialties [91,94]. Differences in personality profile have also been identified between surgical subspecialties [95]. In a Swedish context, Bexelius et al. investigated personality among different medical specialties including surgeons. They found lower ratings among surgeons on agreeableness in comparison to doctors in internal medicine, primary care, and hospital service specialties, and higher ratings on conscientiousness compared to hospital service specialists and psychiatrists [89].

To date, few studies have extensively examined the relationship between the Big Five personality model and surgical outcomes, but some relations have been reported with somewhat conflicting results. Hughes et al. found differences in personality traits between "low" and "non-low" performing residents, where the non-low performing residents rated higher on extraversion, emotional stability, and conscientiousness [82]. In another study, cardiac surgeons showed worse outcomes for surgeons rating in in the highest tertile on openness [96]. However, in a study on laparoscopic simulator performance no significant association was found with personality [97]. Personality may also influence the choice of operative treatment. In a study by Teunis et al. the Octogram Work and Leadership Style Test was used to investigate the influence of personality and the choice of operative treatment. The Octogram Work and Leadership Style Test has been validated and focuses on personality that drive behavior. They found that a higher "pioneer" score correlated with the recommendation of surgical over non-surgical treatment [98]. Apart from technical performance, personality has been associated with burnout, with neuroticism having the strongest association with burnout among several specialties. In recent years, burnout has been found to be prevalent among surgeons [99,100]. For surgical residents, extraversion was less related to burnout [101].

Unsuitability

Unsuitability as a surgeon could include several aspects. As described above, there are variations in both technical and non-technical skills, and the surgical outcome among surgeons varies. For the employer, attrition and remediations could be additional factors of what defines a problematic or successful employee. With remediation of struggling residents, it has been argued that program directors invest 80 % of their time on 20 % of their residents. Even though struggling residents is not uncommon, it does not automatically imply future unsuitability and there are strategies to support their development [102]. While candidates may possess traits that qualify them in certain areas, they may simultaneously exhibit deficiencies in others. It is therefore important to use broad assessments and focus not only on top-achievers in one area [103].

Unsuitability among surgical residents in Sweden has been studied by Hagelsteen et al. In a survey of experienced surgeons, 85 % reported having encountered a resident whom they considered unsuitable as a surgeon. The most commonly cited reasons regarded technical ability, communication and interpersonal skills, and personality, personal resources and skills. As part of that study, in-depth interviews were conducted with 13 experienced surgeons that included the topic of unsuitability. From the interviews, 40 potential warning signs were identified and then categorized into 11 problem domains. The findings formed the basis for an interview guide to potentially be used for assessment of unsuitable tendencies [29]. From the study on potential warning signs, the development of the longitudinal study that this thesis is part of was developed. With the lack of data from a Swedish context it aims to explore both suitability and unsuitability on both technical and non-technical aspects and investigate the development over time.

Feedback

During residency there is a multitude of competencies to learn. Learning takes place in the workplace, often based on feedback from colleagues, as well as from formal advisors and supervisors. It is therefore of interest how an individual responds to and utilises feedback for the purpose of future development. Feedback is imperative for learning and there are several recommended methods that could be employed [104-106]. Even though the origins of competency-based medical education date back over 60 years, the educational system has seen a marked increase in its utilization over the past 20 years. This transition has precipitated a shift in focus from time-based criteria to an evaluation of competencies, accompanied by a more flexible timeframe. The present model for residency training is predicated on formative feedback for the resident from multiple colleagues on all competencies and abilities [107]. Feedback can be categorized into several distinct types. Within the domain of competency-based medical education, formative feedback assumes a particularly significant role [108]. Formative feedback is provided for the purpose of enhancement, while summative feedback is utilised for evaluation at the end of learning [109]. In order to be effective, feedback should be delivered with structure and be based on observations from a situation. During residency training, there exists a range of valid instruments that are utilised for the purpose of facilitating and sustaining structured feedback. Examples of such instruments include Case Based Discussion and Mini Clinical Evaluation Exercises [110]. Whilst the increased evaluations aim to promote development, the perception of an increased assessment burden has been identified as a challenge in the competency-based medical education [111,112].

A number of barriers currently exists regarding feedback, as evidenced by previous research that has indicated an absence of structured, consistent, and timely feedback. This phenomenon has been described in several countries, including Sweden [113-115].

Furthermore, studies have demonstrated conflicting perspectives on the quality and frequency of the feedback provided, with surgeons expressing greater satisfaction compared to the residents [115-117]. The paucity of the learner's capacity to recognise feedback has been posited as a contributing factor. Another barrier that has been described is the fear of criticism, which has been shown to hinder feedback-seeking [114]. Whilst the rationality of fear of criticism is acknowledged, it is asserted that both positive and negative feedback can be utilised for the purpose of effective development [118].

Summary and aims

This thesis focuses on recruitment and selection of surgical residents as a step to improve patient safety. By using validated recruitment and selection methods, attritions rates, cost, and adverse events could potentially be reduced. The currently used selection process in Sweden lacks guidelines, and both technical and non-technical skills as well as the potential for development, are important as a surgeon. There is a lack of knowledge regarding both the selection process and potential assessment methods in a Swedish context.

This dissertation project aims to investigate assessment methods that could possibly be applicable in a Swedish resident selection context. The overarching aim is to evaluate quantitative and qualitative assessment instruments to investigate both technical and non-technical skills, as well as key developmental factors including the handling of feedback, among surgical residents at the time of recruitment. The specific aims for each paper were as follows:

I. Investigate and compare selection processes to surgical residency training in the European Union (EU).

II. Investigate the impact of demographic factors, manual dexterity and visuospatial ability on the performance in two different laparoscopic surgical simulators amongst applicants to surgical resident positions and 1st year residents.

III. Investigate the relationship between personality and ratings on 360-degree evaluations among surgical residency applicants.

IV. Explore experiences and challenges in handling feedback and criticism among early career surgeons seeking surgical residencies.

Methods and material

Study design and participants

General information of the longitudinal study

This thesis encompasses both quantitative and qualitative methods. For assessment methods in a Swedish setting, this thesis is part of a longitudinal study that aims to follow and assess surgical residents repeatedly throughout their residency until they become certified surgeons. The evaluation process during the longitudinal study also encompasses both technical and non-technical assessments of the participants. The technical assessment comprises two different laparoscopic simulators, a manual dexterity test, and a visuospatial ability test. Non-technical skills are assessed using personality ratings, multisource coworker feedback assessments, and through semi-structured interviews.

At the time of enrolment into the study, participants were asked to provide information regarding their sex, age, dominant hand, year of certification, education, previous work experience, surgical experience for open and laparoscopic procedures, laparoscopic simulator experience, current employment status and hospital, and self-assessed 3D vision. The collection of data was facilitated by Research Electronic Data Capture tools (REDCap) [119,120], which is hosted by Lund University. This thesis has focused on data from the point of recruitment and has used the same cohort in Paper II- IV.

The inclusion took place between 2017-2021. Prospective candidates for a surgical residency or locum position in the domain of general surgery, urology, and paediatric surgery were recruited from seven hospital within the South Swedish Healthcare Region and from Uppsala University hospital. All departments involved in the study agreed to participate and provided contact information for applicants during the interview or hiring process. Informed written consent was obtained from all participants. On-site evaluations and interviews were conducted at the Practicum Clinical Skills Centre, Lund, Sweden, a prominent clinical training facility in Europe. The facility is accredited by the Network of Accredited Clinical Skills Centres in Europe (NASCE). Due to the exploratory and multifaceted nature of this study, as well as the relatively low probability of affecting the inclusion, a power calculation was not performed.

Paper I

Paper I investigated selection processes to surgical residency training in the European Union. In addition to a broader study of selection procedures in EU member-states, an in-depth comparative case study of a decentralized and centralized country was conducted. Sweden was selected as an archetypical case of a decentralized selection. The data for this study was obtained through several investigating methods, including literature searches, knowledge within the research group, and analysis of previously conducted semi-structured interviews with 13 experienced surgeons [29]. Ireland was selected as the archetype of a highly centralized and formalized selection process. To obtain comprehensive knowledge, members of the research team conducted two study visits (2018 and 2019) to the Royal College of Surgeons in Ireland (RCSI). During the initial visit, the selection for Core Surgical Training was observed, while the subsequent visit focused on the selection process for Higher Surgical Training. Fieldnotes were taken during the visits and literature on the Irish selection system was reviewed.

For information on the selection processes of surgical residents in the other European Union member-states, searches were conducted on Google and PubMed. The search strategy involved the use of keywords, including the name of all EU member-states, along with the terms “surgical”, “trainee”, “selection”, “specialization” or “application”. Google was included due to the paucity of published material on current selection processes, allowing for detection of information from the department involved in the respective country. It is acknowledged that this approach may not encompass the comprehensive and detailed information about all participating countries. The objective was to provide an overview of the current status of the selection processes. Additionally, an email was sent to 16 board delegates within the Section of Surgery of the Union Européenne des Médecins Specialistes (UEMS) inquiring about their respective selection process. A single reminder was sent.

The selections processes were subsequently categorized into three distinct classifications: centralized, hybrid, and decentralized. The centralized selection process was defined as having a national application, national selection, and national admission/employment process. Hybrid was defined as having one or two of the three, and decentralized as having none (Figure 2).

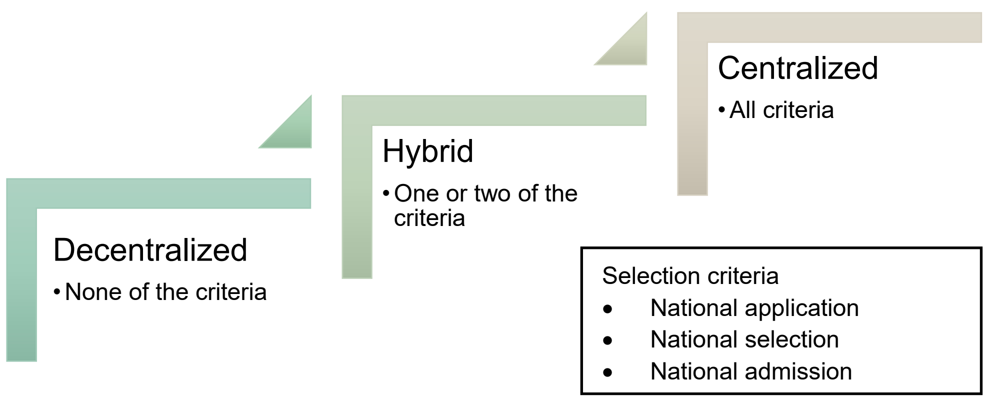


Figure 2.The definition used of a centralized, hybrid, and decentralized selection process.

Paper II

Paper II investigated correlations between background factors and performance in two different laparoscopic simulators. Participants were instructed to complete a background form which requested information regarding their sex, age, dominant hand, year of certification, education, previous work experience, surgical experience for open and laparoscopic procedures, laparoscopic simulator experience, current employment status and hospital, and self-assessed 3D vision. The data was collected using Research Electronic Data Capture (REDCap) tools hosted by Lund University [119,120]. Participants were subjected to a series of evaluations that encompassed the assessment of their VSA, manual dexterity, and laparoscopic simulator performance. Two distinct laparoscopic simulators were utilized.

Visuospatial ability

The VZ-3 Surface Development test [121] is a test designed to assess spatial ability. In this test, participants are instructed to visualize the folding of a flat shape into an object (Figure 3). Participants were provided with an individualized, coded link to a computer-based version of the VZ-3 Surface Development test. The test comprised 12 tasks with a maximum time of one minute per task. The maximum attainable score was 60 points.

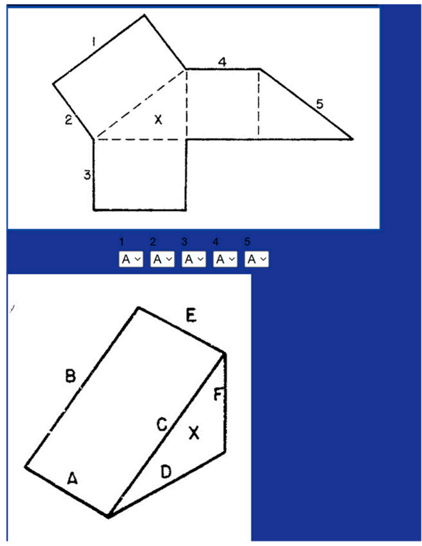


Figure 3. Example of a task from the VZ-3 Surface Development test.

Purdue Pegboard

The Purdue Pegboard (Figure 4) is a standardized instrument employed to assess manual dexterity [122]. Instructions were obtained from the manufacturer and translated into Swedish. Participants were instructed to perform three rounds of the test. The median score of the assembly task was used as a measure of dexterity.

Figure 4. Purdue Pegboard. Photo by Kristine Hagelsteen.

LapSim® (Surgical Science, Gothenburg, Sweden)

LapSim® (Figure 5) is a laparoscopic virtual reality (VR) simulator with 3D display and a haptic feedback platform for force feedback [123]. Participants were instructed to perform the "lift and grasp" exercise from the Ahlberg exam [54]. The time limit was set to 20 minutes, and the task needed to be passed twice within this time frame. The outcome measure employed was a pass/fail system.

Figure 5. LapSim®, a laparoscopic simulator usind 3D vision. Photo by Kristine Hagelsteen.

Simball® Box (Surgical Science, Gothenburg, Sweden)

The Simball® Box (Figure 6) is a 2D hybrid video box trainer that records a video of the performance [124]. Participants were shown an instructional video that demonstrated a double overhand knot and were then permitted to practice with assistance and guidance from an experienced surgeon. Following the initial demonstration, a subsequent repetition of the instructional video was presented, this time without the provision of instructions. The time limit for the task was set to five minutes. The video-recordings were subjected to independent analysis by three experienced surgeons. These surgeons assessed the video recordings according to a previously validated score [125]. Some modifications were made according to the given task (appendix 1, Paper II). The outcome was designated as pass/fail. Discrepancies among assessors were resolved by renewed analysis by the group.

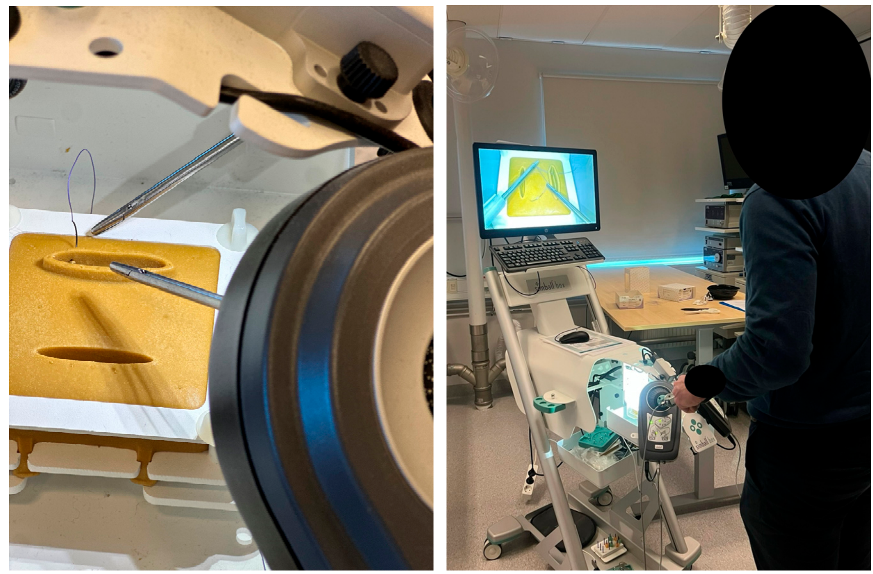


Figure 6. Simball® Box, a laparoscopic simulator using 2D vision. Photo by Kristine Hagelsteen.

Paper III

Paper III investigated the relationship between personality and 360-degree evaluations. The participants conducted personality ratings, 360-degree evaluations, and tasks in two different laparoscopic simulators. The utilization of the laparoscopic simulators served as an objective metric for the assessment of technical skill.

Personality test

Participants were provided with a personal coded link to complete a Neutralized Big Five Inventory (NB5I). The NB5I has demonstrated robust convergent validity with other personality inventories, and is neutralized to minimize the influence of social desirability [126]. The presence of social desirability can result in participants providing responses that align more closely with their desired image than their authentic feelings. The neutralized inventory is designed to mitigate the prominence of the socially desirable option. The inventory consists of 120 items, equally divided between the five factors of the Big Five model. Each item is rated on a five-point Likert scale, providing a quantitative metric of participants responses.

Multisource 360-degree assessment

The Sheffield Peer Review Assessments Tool [127] was utilized in a translated version that is commonly used by the Directors of Postgraduate Medical Education at Skåne University Hospital, Sweden. The assessment tool comprises 25 questions that are

rated on a six-point Likert scale and divided into five categories: good clinical care, maintaining good medical practice, relationships with patients, working with colleagues, and teaching and training. The instrument also includes an option for "not observed". Participants were instructed to share a link to the assessment with six colleagues from diverse professional background. All participants were provided with a personal code for identification. Aggregating the ratings from all coworkers resulted in the creation of variables. Within the overarching category of maintaining good medical practice, subcategories such as "technical skill" and "stress handling" are included. These subcategories subsequently formed additional categories due to their divergence from the other subcategories. The data collection process was facilitated by the REDCap electronic data collection tool.

Paper IV

Paper IV was a qualitative study that explored experiences and challenges in handling of feedback. The participants were interviewed using a semi-structured interview guide that has been previously published. This interview-guide was based on 40 previously identified warning signs [29]. The interviews were carried out by one or two experienced sociologists, with assistance from one or more experienced surgeon in most of the interviews. The participants were de-identified using a personal study ID and they were asked to describe situations without mentioning of identifiable markers such as names. The interviews were audio recorded and stored on external hard drives and at LUSEC, a platform hosted by Lund University that fulfils all demands of safe storage. Verbatim transcriptions were generated. The resulting files were imported into and subsequently coded using Nvivo (Lumivero, Version 14). Two researchers independently coded the transcripts using deductive thematic analysis to the previously identified warning signs. A common coding document was created and discussed until consensus was reached. From the codes themes focused on feedback and criticism were created and repeatedly discussed within the research group.

A variety of coding methods were discussed before choosing thematic analysis. Suggestions encompassed for example an open coding procedure and the prevailing deductive thematic analysis. Thematic analysis was chosen due to the preexisting warning signs that related to the aim [128].

Qualitative research differs from quantitative research, as do the techniques to ensure quality. Common quality criteria in qualitative research includes credibility, transferability, dependability, and confirmability [129]. In this study, credibility was enhanced by two researchers conducting the coding and subsequent discussion of the results until mutual agreement was reached. Further, for additional triangulation the interviews were read and coding discussed with the other authors. All researchers have different backgrounds, thereby bringing different perspectives to

the discussions. To enhance transferability, we described the selection process, the context of the interviews, and the participants to provide readers with information enough to understand the finding and assess relevance. For dependability, ensuring saturation is one of the techniques. Saturation in qualitative research is when no new themes can be identified in the material and can consequently not be calculated before inclusion or the analysis of the material. It is therefore difficult to reach, and also to know when it has been reached [130]. However, the inclusion of 50 interviews at least increases the chance. The last of the four criteria is confirmability. This was enhanced by the continuous discussions of the research process and results obtained.

Qualitative and quantitative analysis

The conducted studies incorporated both qualitative and quantitative analysis. Paper I and IV employed a descriptive and qualitative approach, while papers II-III utilized statistical analyses.

Paper I entailed analysis of previously conducted interviews with experienced surgeons, with the objective of acquiring knowledge regarding the prevailing selection process. Fieldnotes from each study visit to Ireland were pooled and analysed after each visit to enhance understanding of the Irish selection process.

For **Paper II** logistic and linear regressions were employed, depending on whether the outcome was binary or continuous. Both univariable and multivariable analyses were conducted. Due to the limited number of observations, only variables with *p*-values below 0.3 were included in the multivariable analysis. This *p*-value threshold was determined following consultation with a statistician to minimize the probability of missing a significant effect. Odds ratio and beta coefficients were presented with 95 % confidence interval. The odds ratio was utilized to investigate the association between LapSim® and Simball® Box. The analyses were conducted in collaboration with a statistician, employing IBM SPSS Statistics for Windows, Version 25.0.

In **Paper III,** Pearson correlations were utilized to assess the relationship between the 360-degree evaluations, the personality traits, and measure of technical skill levels. Student´s t-test was employed to compare personality ratings, utilizing the mean ratings of the participants and norm values. A *p*-value less than 0.05 was considered statistically significant. Cohen´s d was utilized as a metric to ascertain effect size. Skewness and kurtosis were investigated for all variables.

In **Paper IV,** the transcripts were subjected to a cross-sectional thematic analysis using previously identified warning signs. Subsequently, a common coding document was created and discussed by the authors until consensus was reached before formation of themes.

Ethical considerations

This thesis is part of a longitudinal study that follows surgical residents throughout their surgical training. Ethical approval was obtained from the Swedish Ethical Review Agency (approval number 2016/1050, with amendments 2022-04361-02, and 2023-04850-02). Informed written consent was obtained from all participants. Participants were assigned a unique study identification code, which was utilized during data collection to maintain anonymity. The participants were instructed to refrain from mentioning cities or names during the interviews, and the transcribers were instructed to remove any instances of this type of mention. Despite these measures, both the cohort and the surgical community in Southern Sweden are rather small, which increases the likelihood of identification. To mitigate this potential risk, the primary coding of the interviews was conducted by two researchers who are not surgeons. Additionally, the tests were not administered within the context of a true employment scenario, and the results were not communicated to the employer, thereby mitigating the likelihood of adverse consequences for the participants. All results were presented at group level.

This thesis has been produced with the assistance of Artificial Intelligence. DeepL Write Pro was used to correct and improve the language and readability in Paper IV and in writing this thesis summary. After the use of DeepL Write Pro the author has reviewed the text and takes full responsibility for the content.

Results

Paper I

Selection processes in the European Union

Selection processes in 27 countries were identified and analysed. The selection processes were divided into a scale of centralized, hybrid, and decentralized (Figure 7).

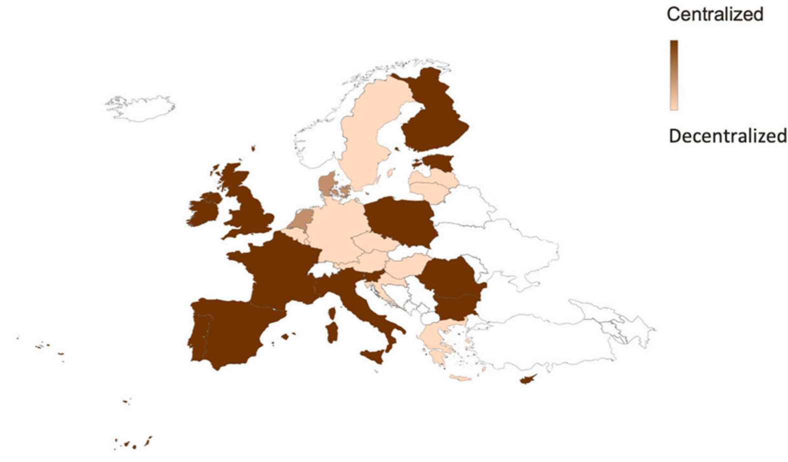


Figure 7. The selection processes to surgical residency programs in the Eroupean Union organized in a cenralized, hybrid, or decentralized fashion.

Seven of the 16 approached UEMS Board Delegates responded. One country was excluded due to lack of information. There was a wide range of selection methods, from strict national and centralized, to a decentralized system working at hospital level. None were identical. Transparent and standardized selection with clear weighting of different criteria was used in a few countries. Interviews were the most frequently used method, but the type varied from ad hoc to Multiple Mini Interviews (MMI). Eleven countries had a decentralized process.

Of the countries with a centralized selection process, Ireland had the most elaborate process and was the only country to formally test for technical skills. Spain had a

centralized process based on examination results. The conditions differed from one country to another, and in Poland, there were fewer applicants than positions, which resulted in more or less all applicants being accepted. Other countries had a centralized system due to the size of the country with only one teaching hospital.

In the countries with a decentralized selection system, some had a standardized consensus across the university hospitals, and others had local decisions on recruitment. In Greece, positions are advertised on the official website of each local authority for Public Health and recruitments are made on a "first come, first served" basis.

Comparison of a centralized and decentralized selection method

Sweden and Ireland were selected as contrasts on the centralized-decentralized scale. Sweden had a strictly local selection at each hospital with no standardized guidelines. A common system was to utilise a 6-12-month locum position, after which the head of the department decided whether the aspiring surgeon would receive a permanent residency position. The method had fairly strong support, but the lack of transparency and the unstructured evaluation of the aspiring surgeon was considered problematic, especially by those who had experienced unsuitable candidates slipping through the system.

In Ireland the selection criteria is described in detail on the RCSI website [131]. In contrast to Sweden, surgical training in Ireland takes place in two stages, with a primary selection at the beginning of surgical training and an additional selection process after two years. The first selection, for CST, was included for the comparison in this study. Approximately 100 candidates applied for 60 positions and were tested on both technical and non-technical skills. Four MMIs with two assessors were used, focusing on different professional areas. The assessment was done individually by the assessors using a Likert scale. Prior to testing, the assessors were instructed on how to conduct the assessments in a joint briefing session. The RCSI used a metric system for scoring, with a maximum of achieving 100 points. Overall, there was strong support for the system from the assessors, but concerns were raised about the possibility of identifying unsuitability during the brief interviews.

Participants Paper II-IV

Seventy-three applicants were invited to participate, 53 accepted, and 50 completed. The most common reason for not participating was non-response (n=10). Most of those not included were later not offered a position as a locum or resident. The median age of those included was 30 years (range 27-41) and 60% were men. Seven were not offered employment. All but one had experience from assisting in laparoscopic procedures, ranging from "never" to "> 100 times". One applicant was a specialist in another medical field. Background variables are shown in Table 2.

Table 2. Background characteristics for participants in studies ii-iv.

Background variables		Count
Sex	Female	20
	Male	30
Age, median and range (years)	30 (27-41)	
Country of medical education	Sweden	31
	Other Nordic country	6
	Europe	13
Years after medical license, median and range	2 (0-16)	
Specialty	General surgery	34
	Urology	15
	Pediatric surgery	1
Previous specialist in another field	No	49
	Yes	1
Self-assessment: "I can see 3D on movies"	No	0
	Yes	50
Gaming experience	No	25
	Female	15
	Male	10
	Yes	25
	Female	5
	Male	20
Dominant hand	Right	44
	Left	4
	Both	2
Work status	Trainee	34
	Locum	9
	No employment offered	7
Laparoscopic simulator training experience	No	24
	Women	14
	Men	10
	Yes	26
	Women	6
	Men	20
Assisted in laparoscopic procedures	Never	1
	1-10 times	14
	11-20 times	15
	21-50 times	14
	51-80 times	1
	81-100 times	2
	>100 times	3
Independent laparoscopic experience	Never	34
	1-10 times	11
	11-20 times	1
	21-50 times	3
	51-80 times	1
	81-100 times	0
	>100 times	0

Paper II

All participants completed the LapSim® and Purdue Pegboard tests. One video recording from Simball® Box and three results from VZ-3 Surface Development test were missing due to technical failure. There was an initial disagreement among the raters on 20 of the 49 Simball® Box video recordings before consensus was reached.

Some 13 participants passed the LapSim® test, 16 passed the test on Simball® Box test, and five participants passed both exercises. The odds ratio between passing the LapSim® test and Simball® Box was not significant. There was a significant association between the VZ-3 Surface Development test and Simball® Box (95 % CI: 1.028-1.2, p-value < 0.01). This remained significant after multivariable analysis. The odds of completing the task increased by 11 % with each point gained on the VZ-3 Surface Development test. No significant associations were found with background factors such as gaming experience, or manual dexterity (Table 3).

Table 3. Regression model between pass/fail results from Simball® Box, included background factors, and tests.

	OR (95 % CI)	p-value
Univariable analysis		
Sex (male vs female)	2.50 (0.67, 9.39)	0.175
Simulator experience	0.94 (0.29, 3.11)	0.921
Independent laparoscopic experience	0.91 (0.25, 3.27)	0.884
Gaming experience	2.26 (0.66, 7.70)	0.192
Visuospatial ability	1.11 (1.03, 1.20)	0.008*
Pegboard ”assembly”	1.03 (0.93, 1.13)	0.627
Multivariable analysis		
Sex	3.13 (0.55, 17.97)	0.200
Gaming experience	0.93 (0.17, 5.00)	0.929
Visuospatial ability	1.11 (1.03, 1.21)	0.010*

Variables with p-values below 0.3 are included in the multivariable analysis. $^{*}p < 0.05$.

There was no significant association between sex or gaming experience and the VZ-3 Surface Development test score. For the univariate analysis of the continuous outcome on LapSim®, only weak associations were found, which disappeared in the multivariate analysis. There was no association between the VZ-3 Surface Development test and LapSim® (Table 4).

Table 4. Regression models for binary and continuous outcome on LapSim® and included background factors and tests.

	OR (95 % CI)	*p*-value
Binary outcome		
Univariable analysis		
Sex (male vs female)	1.71 (0.45, 6.58)	0.432
Simulator experience	1.11 (0.31, 3.92)	0.877
Independent laparoscopic experience	1.48 (0.39, 5.54)	0.563
Gaming experience	1.88 (0.52, 6.85)	0.337
Visuospatial ability	1.00 (0.95, 1.06)	0.897
	Beta-coefficient (95% CI)	***p*-value**
Continuous outcome		
Univariable analysis		
Sex	9.68 (0.04, 19.31)	**0.049***
Simulator experience	3.80 (-5.98, 13.58)	0.439
Independent laparoscopic experience	-0.49 (-11.03, 10.06)	0.927
Gaming experience	9.62 (0.19, 19.05	**0.046***
Visuospatial ability	0.30 (-0.13, 0.72)	0.166
Multivariable analysis		
Sex	8.25 (-3.09, 19.59)	0.150
Gaming experience	4.64 (-6.66, 15.94)	0.412
Visuospatial ability	0.25 (-0.17, 0.67)	0.229

Variables with *p*-values below 0.3 are included in the multivariable analysis. *p < 0.05.

Paper III

A total of 38 participants were included in the analysis. One video recording from the Simball® Box was missing due to technical failure, and two participants did not complete the personality ratings. Twelve participants received zero 360-degree evaluations (range 0-6). Those with at least one rating were included in the analysis.

Compared to the norm (norm value = 2.0), the group rated higher on four personality domains: agreeableness, conscientiousness, emotional stability, and openness (Table 5). Agreeableness, $t(47) = 4.62$, $p < 0.001$, $d = 0.67$), conscientiousness, $t(47) = 6.44$, $p < 0.001$, $d = 0.92$), emotional stability, $t(47) = 6.79$, $p < 0.001$, $d = 0.98$), and openness, $t(47) = 2.27$, $p = 0.033$, $d = 0.33$).

Table 5. Personality ratings from the Neutralized Big-Five Inventory.

Personality traits	Mean	SD
Extraversion	2.01	0.56
Agreeableness	2.23	0.35
Conscientiousness	2.34	0.38
Emotional stability	2.50	0.51
Openness	2.18	0.54

Norm values for personality traits = 2.

For the 360-degree evaluations, the ratings were in general high, with the mean responses between 4.79-5.17 on a 6 grade Likert scale (Table 6).

Table 6. Results from the 360-degree evaluations using a Likert scale from 1-6.

360-degree evaluations	Mean	SD
Clinical competence	4.84	0.56
Medical standards	4.79	0.59
Patient relations	5.17	0.54
Colleague relations	5.05	0.52
Stress handling	5.03	0.79
Technical skill	4.87	0.64

There were six correlations between the 360-degree ratings and personality, four of which were with extraversion. The strongest were between extraversion and maintaining good medical practice and technical skill, and these were the only ones that held up to multiple significance tests. (Table 7). There were no significant correlations to the results from the laparoscopic simulators.

Table 7. Correlations between variables from 360-degree ratings, personality traits, and technical performance from two different laparoscopic simulators: LapSim® and Simball® Box

360-degree ratings	Big Five Personality Dimensions					Technical performance	
	Extraversion	Emotional stability	Openness	Agreeableness	Conscientiousness	LapSim®	Simball® Box
Good clinical care	0.38*	0.24	0.03	-0.20	-0.04	0.02	0.28
Relationship with patients	0.18	0.19	0.10	0.08	-0.12	0.12	0.21
Working with colleagues	0.37*	0.16	-0.03	-0.08	-0.03	0.03	0.18
Maintaining good medical practice	0.56*	0.30	-0.20	-0.35*	0.08	0.04	0.18
Stress handling	0.32	0.25	-0.08	-0.31	0.02	-0.09	0.14
Technical skill	0.67*	0.23	-0.39*	-0.27	0.27	-0.05	0.05

. * = p <0.05

The "teaching and training" category was excluded due to low reliability. Exploratory regression tests were conducted to determine whether other variables predicted the 360-degree evaluations after extraversion was excluded. None of these results would be significant after multiple significance tests.

Paper IV

From the identified codes four themes relating to handling of feedback and criticism were constructed: 1) reflection and processing of criticism, 2) emotional response to criticism, 3) deficient feedback culture, and 4) navigating criticism in a hierarchical system. Table 8 illustrates examples for each theme.

Table 8. The four themes related to feedback and criticism identified from the interviews with examples of quotes regarding each theme.

Theme	Examples of quotes
Reflection and processing of criticism	"I listen to it, to the whole situation, and then analyse my behavior. If it is an experience that the person has been part of then that experience is true for that person and then I have to apologize for the misunderstanding. If it is something I have done wrong, I usually feel that it was my fault. But sometimes when criticised by patients I have thought it through and discussed it with a supervisor and felt like I should have done it in the same way next time. Yes, one part might have been my fault but not another one, and I have to listen and learn, to get criticised but not let it get to me that I have done something wrong if a colleague can support the medical decision. But it is much about communication, it can be that kind of flaws." (IP 7)
	"I think you should accept all critique, but you don't have to implement it if you don't fell it is reasonable. The times I have been criticized I have listened to it but it is not always you find it fair, perhaps I was stressed or had other reasons for my actions. But the times you feel it is constructive you have to be open for it, we are not perfect." (IP 11)
Emotional response to criticism	"Or colleagues, it can also be colleagues that yells at you for quite bizarre reasons, when you don't really understand. Then I can doubt what I am doing and doubt myself." (IP 30)
	"Yes, it is hard to say, I take it with me. I might remember things a rather long time, my mom always nags on me for this. I am usually good at remembering the negative and bad at remembering the positive. It depends on what it is, of course it is never fun to receive negative criticism but at the same time you have to take it and make something good of it." (IP 14)
Cautious feedback culture	"How do I do, good question. I think I in general, as everyone else, think it is hard to provide critique. But I usually say, in this situation I saw that you did this, and what you can think about in the future is to try this other way." (IP 1)
	"I don't like to step on anyone's toes, but if I think something is wrong, I say it sometimes, but not always. I try to be rather diplomatic." (IP 42)
Navigating criticism in a hierarchical system	"It depends on, unfortunately it is not easy, if it is a senior doctor higher in the hierarchy, to criticize them, at least not at our clinic. So it depends on how susceptible that person is and how urgent it is. If it is something dangerous of course I will act, but if it has something to do with that there are different styles, styles of treatment, and we don't agree with each other then it is just an opinion, and it depends. But unfortunately, it is not that easy and even if it is constructive criticism to someone senior, not everyone is susceptible, all though some are." (IP 31)
	"But like I said before, because there are hierarchical structures in healthcare, it can be hard to discuss with a consultant. If it is pure discrimination and things that happen between colleagues then I try to bring it up with a chief or question in the ongoing situation. It depends on what has happened, I have for example when I have felt suppressed by a consultant emailed him. And I told him how I felt and that I thought it is important with respect. Interviewer: Did you get a good response? IP: No, something like have a good life. But it still felt right, I think it is important to stand your ground." (IP 13)

Reflection and processing of criticism

Reflect and analyze criticism before accepting it was commonly described by the participants. This process could also lead to rejection of the critique if it was found to be unreasonable.

"I usually don´t just accept the feedback as it is, because even if people are usually right, but not always, you have to think a little bit to integrate into your own world" (IP15).

Reasons for the criticism to be deemed unreasonable included for example when it was given in a stressful situation, if it regarded preferences for different methods, or if it was provided in a judgmental manner. Some participants stated an openness for criticism and a willingness to improve.

Emotional response to criticism

Some of the participants described difficulties handling critique as well as an overall sensitivity that could lead to self-doubt. Differences in the response to critique varied, from trying to improve to approach similar situation with more caution thereafter.

The emotional response to criticism was found to vary widely, ranging from irritation to indifference, and even to the development of self-doubt.

"If it is given in a very condescending way, very patronizing and mean, without any reason that could help me move forward, then I get irritated rather than accept it" (IP 38).

The surrounding situation could also impact on the perception of unfair criticism. One participant described a particular instance in which they had attempted to solicit assistance and had been rebuffed on the grounds of time constraints. Subsequently, they were a few minutes later subjected to censure in a meeting for their handling of the same scenario.

"at that moment you feel really bad, like why did I become a doctor? (...) You might not always think of how you provide criticism, and sure, you can understand that perhaps you don´t have the energy to sugarcoat it and instead it gets very straight. On a good day I can handle it, but on a bad day it isn´t good when I get home" (IP 4)

An environment that included the presence of unfair criticism lead to additional difficulties. Examples of situations of unfair criticism included being criticized for something that had not happened or being shouted at for unfair reasons. This form

of unwarranted criticism was even described to potentially affect career choice, as one participant left research due to its impact.

"I had the opportunity to visit another group, and I saw that they had this great equipment, and I hadn´t had any results, every measurement was different. I told the research group leader, who scolded me for not finding my own solutions and being more creative, and for always relying on money. It was upset because I did not expect to be yelled at in front of the whole research group, (...) and then I quit after many years of research" (IP 9)

Cautious feedback culture

A learning environment with a cautiousness or unwillingness to provide criticism was described by several participants. This cautiousness concerned both feedback for themselves and colleagues they were supervising.

One of the participants stated to have not received criticism during a locum position, and thereby not been able to correct the behavior that led to not being hired as a resident. Some participants who had experiences from working abroad felt that the Swedish work culture was more cautious regarding feedback and criticism in comparison to the other European countries that they had worked in. Those participants further described an absence of feedback in Sweden, and that feedback could not be taken for granted but instead the responsibility was on the junior doctors to ask for it.

"In my experience, you don´t like to criticize others here, and you try not to. You might not like it, but you don´t say anything. I would say it happens much more in [country] that you get feedback, often on things that were not good, and you have to deal with it because it is not always in a way that is easy to receive. But I think you get more feedback and a better understanding of what you did right and wrong, and here [in Sweden] you have to ask. Did I do a good job, or could you give me some feedback so I know? What can I do better?" (IP 17).

To provide feedback to others was also described as difficult by several participants. Positive feedback and providing suggestions for improvements were regarded easier than providing negative feedback. Some expressed a fear of conflict as the reason for avoiding providing criticism and to avoid taking fights. This behaviour persisted even if the fights were thought to be adequate.

Navigating criticism in a hierarchical system

The participants experienced a hierarchical environment with challenges regarding criticism and feedback. It was reported by several participants that the approach

adopted varied according to whether the communication was directed towards a subordinate or a superior.

"...if it is someone superior to me doing something wrong, I think I just ignore that person and do what I think is right and change without them noticing. But if it is someone lower in rank it is often due to inexperience and then I might be more verbal in telling them what needs to change" (IP 24).

Both respect of hierarchical structures and fear of retaliation were reasons for the behavior. A few additionally confessed to rather than engaging in discussion with the superior, they chose to correct the issue in secret. In some cases, both the hierarchical structure and the severity of the error were considered:

"It is a conflict; on the one hand there is the element of fear, and on the other hand there is the seriousness of the mistake. If the mistake is serious, I have to overcome my fear. You don´t criticize your boss for a minor event, but if it´s someone you´re teaching you can give them feedback" (IP15).

A commonly expressed solution in situations relating to superiors was to use questions instead of potential critique. This approach was adopted by the participants to allude to a paucity of knowledge or to be able to use examples of how they have observed others act or conduct a procedure. This behavior manifested the educational role towards the supervisor. Conversely, when addressing an individual in a supervisory capacity, suggestions for enhancement were employed.

There was an experienced difference in regarding if the criticism regarded technical or medical situations in contrast to behavior, which was perceived as more challenging to comment upon.

"...behavior is much more sensitive. I don´t know what I would do if an older colleague behaved really badly towards someone. That would be a really difficult situation" (IP 40).

Discussion

This thesis has evaluated both technical and non-technical aspects around time of recruitment of surgical residents in Sweden. Recruitment of surgical residents is a complex and multifaceted field. A competent surgeon needs to master several technical and non-technical skills, and future job performance is challenging to predict. Further complicating the issue are the different prerequisites depending on country and recruitment opportunities. By increasing the knowledge of a Swedish cohort of applicants and their baseline values, these results could be the first step in evaluating and improving the selection process in Sweden. This could lead to a reduction in the likelihood of chance playing a role in selection and that unsuitable candidates are "slipping" through the system. Possible improvements for recruitment and the learning environment are suggested, as well as areas for future research.

Recruitment process and prediction of job performance

Paper I shows great variation in the recruitment processes of surgical residents in the European Union. A majority of the countries used non-MMI job interviews or academic examination to select candidates, which have relatively low predictivity of performance [41,42]. Although possessing the strongest evidence, a structured interview was not often used.

By comparing a decentralized selection process, such as Sweden and Ireland, both disadvantages and advantages were identified and the differences between countries became apparent. Having a national, structured, and centralized selection increases transparency and have the potential to facilitate an evaluation of the process. The decentralized process has the main advantage of being flexible and adaptable to local needs [132]. A main difference between the countries is the use of locum positions after recruitment, which are in part used as an extended selection process in Sweden. The Swedish local recruitment relies heavy on previous job performance and was considered adequate by the experienced surgeons interviewed in a previous study [29]. This is in line with previous research that have shown that previous job performance is a valid assessment basis if the period of work contains assessment and feedback in a structured manner [23,133]. However, the structured assessments and feedback seem to be lacking according to the early career doctors in Paper IV, and

that would make the current selection process of a work trial period of questionable validity. If future employment status depends on assessment during the time as a locum, structured evaluations must be done, and constructive feedback must be provided. This is especially important in the case where the employer does not intend to extend the employment period into a permanent position. Otherwise, the invested months on training for both the employer and the new doctors might be of limited value. Instead, with the opportunity for feedback that could lead to correction and improvement during this time the potential to maintain a candidate would increase for the employer. The evaluation of the performance and guidance for improvement on professional development would be valuable for the new doctors as well, whether the position is prolonged or not. The current feedback culture and communicated expectations needs to improve both for the current selection process based on work-trial periods, and for the possibility for the new doctors to perform according to anticipation.

The large national testing facility and high-stake exam day/days used in Ireland have several advantages, such as standardization and fair recruitment. However, the implementation of a national testing would come with increased costs that currently would not be justified in Sweden given for example the small number of applicants and decentralized setting. Sweden has different prerequisites with locum positions and no structure for national control over the current positions. However, there is still room for improvement while maintaining the decentralized process and making it more evidence-based. Firstly, to advertise externally and to ensure that there are several candidates to assess for a position. However, it has been shown that “home” candidates receive higher interview ratings [134], and interviewers should be attentive to this risk. Secondly, a structured interview with standardized questions should be included in the recruitment process [23,24,33,34,135]. A structured interview correlates with residency performance in the first year [136]. Ideally, more than one person should be involved in the interview [33]. In Ireland, the MMIs are rated according to a metricized system using a Likert scale for evaluation in proximity to the interviews. Although not using MMIs in Sweden, a similar system could potentially be employed by either scoring questions or sections of an interview. By assessing and writing down results on each part as well as standardizing the questions would make comparison of candidates easier and increase transparency of the results. Using a Likert scale to evaluate an interview will still to some extent remain subjective, and in Ireland, both a meeting to align the assessors prior to the interviews and the use of two assessors at each station improves the assessment. If conducted in a Swedish setting, the risk of subjectivity should still be taken into account in similar ways. For additional standardization, structured interviews of the references could also be an improvement. Other potential development could be the to use MMIs or Situational Judgement Tests (SJT). However, MMI would be costly in the current decentralized system [137]. SJT would also bring additional costs, but could be used for initial screening for candidates, before deciding on who to interview [138]. However, with the currently small number of applicants in a Swedish context, an

additional screening process would in most cases be of questionable additional value. For this reason, the resources that would be required to add SJT into a selection process would likely not be a good investment.

Virtual interviews could be an option and have been used more after the covid-19 pandemic. In the USA, the application process is costly for both applicants and employers [46], but with the generally relatively short distances in Sweden, the costs are not comparable to the USA. The advantages would therefore not be as clear in a Swedish setting.

Most countries use a selection system designated to find top candidates, but the opposite, identifying those who are unsuitable, may be even more important. Up to 30 % of residents require remediations, and this is both costly [4,22,139-141] as well as affect patient safety. Attrition adds additional costs [4] and delays the time to specialist certification for the individual physician. However, the research on costs is not conducted in a Swedish setting and is therefore not transferable because of the different selection processes with for example screening of more applicants. In Sweden, there is no information of the number of residents at any given time. The only information available is the number of doctors who get certified specialists each year. To the author's knowledge, there is no research on remediation or attrition either, but the problem of unsuitability has been established [29]. Additional challenges with attrition rates arise from the current employment system with locum positions where there is not always an intent to renew, as with short-term employment during for example parental leaves. Attrition rates, possible reasons for attrition, and costs of recruitment and remediation should to be investigated in future research.

Technical skill

In our study, only Ireland used structured testing of physical technical skills during the study period in Paper I and have since discontinued [142]. Their process for recruitment is now conducted virtually, still including test of psychomotor skills, VSA, and perception [143]. Although the use of laparoscopic simulators for technical testing would be costly and time consuming, the assessment of technical skills prior to or at the beginning of employment is appealing. Studies have shown that some individuals seem unable to achieve proficiency despite dedicated training in laparoscopy [50-53], and both time and resources could be saved if early assessment is used for formative feedback. However, the use of laparoscopic simulators as a summative high-stakes exam at recruitment has several drawbacks, such as that only baseline values are used, that technical skills can be learned, and that it is resource demanding [144]. Instead, as seen in Paper II, VSA seems promising as a proxy, which is in line with previous research [71]. This could be a promising method for a selection

in Sweden as well, with low cost and feasible to perform at every hospital. However, there are conflicting results as to whether VSA is trainable [68,69], and a variety of different tests have been used in previous studies [70]. For further investigation, the focus should be on trainability and to investigate which test, or tests, that are most related to laparoscopic performance. In a Swedish setting with locum positions and the opportunity to try different subspecialties before choosing one, another way to improve the current situation could possibly be to implement a mandatory training program on laparoscopic simulators early in training. This would increase the structured evaluations while still maintaining the current process of locum positions. However, the manner in which such training might be implemented lies beyond the scope of this thesis.

We found a significant association between performance on the Simball® Box and visuospatial ability. No other associations were found between performance on either simulator and background factors, VSA, or dexterity. The lack of association between VSA and the results from LapSim® scores could possibly be explained by the 3D setting, which presents visuospatial depth and may compensate for lower VSA. No differences in learning curves between VR simulators and box trainers have been shown [145]. For additional evaluation of the potential use of VSA, learning curves should be studied. If limited VSA results in difficulty or inability to achieve proficiency, the testing of VSA as part of a selection process is strengthened. It could also be used to identify residents who are in need of additional support as well as to individualize training programs.

No significant correlations with sex or gaming experience were found, despite that other studies have demonstrated such correlations [65,76]. A problem with video gaming in Paper II is dichotomization. Binary variables were used, with the possibility of large differences in hours played under the "yes" category. Both this issue and the small sample size may explain the lack of significant results. Similarly, this study did not find support for previous laparoscopic training being predictive of laparoscopic performance, which has been reported previously [75] . This may also be due to the binary variables. In addition, when divided into groups based on number of procedures, each group was too small which made the power very low.

Non-technical skill and halo effect

There are several aspects of non-technical skills, and this thesis has focused on personality, 360-degree evaluations, and handling of feedback early in the career among applicants for a surgical residency position. In Paper III extraversion was found to correlate positively to higher ratings on 360-degree evaluations. Extravert people in general can easily connect with other people, and in medical school extraversion has been shown to correlate to positive evaluations during clinical

years. This has not been found in pre-clinical years, perhaps due to increasing amount of social interaction during the clinical years [146]. In Sweden, the 360-degree evaluations are commonly used to assess residents early in surgical training regarding progress and fit. They are sometimes used as an aid in the decision on whether to promote a locum to a residency position.

A problematic area with the assessments was the correlation between extraversion and technical skill on the 360-degree evaluations, together with the lack of significant correlations to the objective measures in the laparoscopic simulators. Whether performance in the laparoscopic simulators used in the present study truly reflects technical skill cannot yet be determined and conclusions must await the results of the ongoing five-year longitudinal study. Further, it is possible that candidates who rate high on extraversion are in fact more technically skilled. Theoretically, they might have been more active and successful in asking for coaching and operating time. However, since the participants are early in their surgical career it is unlikely that this alone would explain the results. Instead, it might be that the higher scores reflect their greater social skills. This is known as a "halo" effect [147]. The reflection of one assessment into another has been demonstrated in several studies, for example in assessments of surgical residents [148], and between attractiveness and academic performance [149]. The halo effect has not been found in all studies [150], and when identified, it could be reduced by group assessments [151]. The results in this study should be interpreted cautiously since only one question in the assessment form regarded technical skill, and the primary focus of the 360-degree evaluations regards non-technical skills. However, with these results, it points to the need for objective assessments and that it is important to use trained assessors.

Considering the problematic feedback culture overall, a new evaluation of 360-degree evaluations after training of assessors should be considered in a Swedish setting. However, with the descriptions of an "unwillingness", down prioritizing or avoidance to give feedback and critique to colleagues in Paper IV, it is possible that there is a need for other types of evaluations in the Swedish setting. Although 360-degree evaluations have demonstrated promising results in other countries[152,153], the cultural differences in directness may have influenced the outcomes, a notion that is further supported by the absence of outliers in our own study. Of course, the lack of significant results could be due to all our included participant in fact being competent. But it could also be due to limitations, e.g., the fact that the evaluations have not been assessed in Sweden. With the results in this study, where all participants were rated in the upper half, we found no support for this approach to select between candidates. Considering that 360-degree evaluations are widely used in residency training today, that extravert persons are rated higher, and the described problematic feedback culture, there is a need for further investigation.

Participants in Paper III rated higher in four of five personality domains. Previous research has shown that orthopaedic surgeons have been able to outsmart this kind of test and answer desirably in line with having “the requested personality”, or a personality suited for the job [154]. However, the NB51 has been designed to minimize social desirability [155]. This study further included applicants who did not receive a position, and the testing was not part of an actual recruitment process, which would additionally decrease the risk. The results are similar to previous studies [91], but comparison is challenging due to cultural differences between countries, and since the working climate has changed over the years [91,92]. With the differences in personality across countries, there is no clear surgical personality to date. If personality inventories are to be used as an aid for recruitment, larger investigations should be conducted and further delineate the subcategories between for example urologists and general surgeons. Additionally, if the personality does correlate to performance would be further of interest, and if so, the underlying reasons.

Assessment, feedback and the learning environment

While feedback is essential for professional development in an educational context, Paper IV revealed both shortcomings and individual challenges in the ability to handle criticism effectively. With the lack of feedback already described in several countries [113-115], the participants in this study who had worked in other European countries stated to experience an even more cautious approach in Sweden. Despite the challenges inherent in precise comparison, this suggests that the Swedish feedback culture may potentially entail a greater number of problematic elements. An additional aspect of this matter derives from the participant who stated to neither having been given any feedback nor been informed during the locum position that the expectations had not been fulfilled. This could potentially mean that both feedback and communication on expectations are lacking. On the positive side, some of the participants expressed a willingness to educate.

Previous studies have shown a discrepancy between educators and learners regarding the perceived frequency and quality of feedback. One potential development that could lead to greater alignment is education and training of assessors [115-117]. The existing system is beset with several challenges regarding the provision of feedback. One such challenge is the fact that residents often work with several different supervisors, and this fragmented scheduling affects the potential for feedback negatively [156].

A process of filtering and analysing criticism was described by the participants in this study. This analysis was conducted prior to the decision of whether or not to accept the criticism. This approach has the potential to be perceived as arrogant or problematic. However, within the context of an existing environment with unfair

criticism, it might instead be a way to survive and thrive. The management of criticism has the potential to influence development and life decisions, as evidenced by the one participant´s decision to discontinue a career in research after being criticized and yelled at. Some of the participants described a sensitivity to critique that could lead to internalization and self-doubt. This finding aligns with previously described barriers for feedback-seeking, where fear of being criticised has been mentioned [114].

The Swedish society is commonly characterised as less hierarchical. However, this study reveals that a significant degree of hierarchy remains at least in some areas in healthcare. The subjects described a fear of retaliation connected to hierarchical relations. In light of a recent study on Swedish doctors, which revealed that 17 % of respondents reported having been punished for providing criticism concerning workplace issues, this concern may be well-founded [157].

From our findings we suggest some improvements:

1. Schedule new doctors to work more frequently with their supervisor.
2. Clearly state expectations regarding competencies during locum positions.
3. Educate and train both residents and supervisors in feedback and assessment to align their capabilities.
4. Encourage and construct a system where feedback is appreciated on all levels across hierarchical positions.
5. Identify young doctors willing to educate others early and train them in different feedback methods to build a stronger feedback culture for the future.

Conclusion

This thesis has investigated the recruitment and selection process as well as technical and non-technical aspects among surgical residents at time of recruitment in Sweden.

The current selection process in Sweden is conducted locally, lacks transparency, and often relies on job performance during work-trial periods. This decentralized selection process differs remarkably in comparison to for example Ireland which has a national, structured selection process with clear guidelines. Throughout the rest of the European Union there is a multitude of different selection processes used and there is no consensus on what should be included. Both decentralized and centralized selection processes hold advantages as well as disadvantages. From a Swedish perspective, potential improvement in the decentralized context should be to ensure multiple candidates and use structured interviews.

For evaluation of technical ability, test of visuospatial ability holds the largest potential to be used as a future selection method due to the correlation to the performance in a 2D laparoscopic simulator. No other background factors correlated to performance in neither the 2D simulator nor the 3D simulator. Knowledge of visuospatial ability could also be used to individualize laparoscopic training programs.

For assessments of non-technical skills 360-degree evaluations are problematic as part of a recruitment process since extravert persons received higher ratings. The higher ratings included ratings on technical skills, but without any correlation to laparoscopic performance. A possible bias might be the ”halo effect”, where strengths in one area reflect into other areas as well. The halo effect should be taken into consideration when performing assessments and highlights the need for objective assessments.

The general feedback culture in Sweden was described as cautious and affected by hierarchy. The participants expressed difficulties in handling of criticism and an environment where unfair criticism was present. To filter and analyse feedback before accepting it was a common strategy, which might be a method to thrive in this environment. Suggestions to improve the overall feedback culture include for example: scheduling new colleagues to work with supervisors regularly, clearly express expectations on competencies during the first employment time, and to train and educate all employees in feedback and assessment.

Overall, the selection process in Sweden could be improved by adopting a more evidence-based approach such as interviewing multiple candidates for a position and to use structured interviews as a start. The current feedback culture has deficiencies, but on the positive side young doctors describe a willingness to educate. Training in feedback and assessment would improve the overall feedback culture, the potential for development during residency, and increase the structure for assessment during locum positions.

Future perspectives

The fields of surgical selection and education have gained an increased interest in the last years with several attempts of improvements, such as the extensive testing in Ireland [27]. Focus on education and individual progress in different technical and non-technical aspects have been investigated, but the questions if the training of a surgeon is for everyone or not, still fascinates and is part of the surgical community's discourse.

However, the recruitment of residents has received less attention in Sweden compared to some other countries. Recruitment is the first step in forming the future surgeons. While there are several improvements that could be done according to the current literature, as mentioned above, there are also several aspects that are still clouded or unexplored to date.

Upcoming changes in medical licensing and internship

Some changes to the selection process are already upcoming. In 2021 there was a change in the Swedish medical school curriculum. Previously, medical school included eleven semesters and a mandatory subsequently clinical work for 18-24 month as an intern before becoming licensed. With the recent change, medical school now includes 12 semesters and the students become licensed after they graduate [158]. Basic training, called Bastjänstgöring (BT), has been implemented as a new form of clinical rotation of 6-12 months after graduation. It can be integrated as part of residency but it can also be performed as separate employment [159]. With this implemented change, the time for residency has increased from at least five to at least 5.5 years [160]. Surgery is no longer a mandatory rotation [161]. Thereby, the work-experience will potentially be reduced by the same amount of time for those applying for a residency position. During work as an intern a mandatory rotation in surgery of at least three months was included [162]. This forms a potential challenge for recruitment since many departments recruit from the cohort of interns. During the internship period, they get to work full time for three months in a surgical department and get experience regarding what a surgical career could be like.

Comparing with only medical school rotations, the internship forms additional insights and possibilities. From the employers´ perspective, the three months of work observation time will disappear. For the early career doctors, those who have not decided on their future speciality during medical school might not get to experience the surgical department and therefore might not consider it at all. With the system recently implemented and no students from Swedish medical schools have yet to have graduated under the new system, the impact these changes will have is still unclear.

Getting the medical license at university means immediately being legally responsible for patients in contrast to the previous employment as an intern, where the responsibility lies on the supervisor. The introduction for a new employee varies largely between departments, and the ongoing reformations could form a need for a better and structured introduction, but also for an improved recruitment. With less experience and one less evaluation before being singlehandedly responsible for patients, some problems might have been concealed. As for the recruitment process, with less working time there will be fewer references and more unfamiliar applicants as well as less potential for evaluation of technical skills before employment. This could strengthen the need for a structured process with objective measures in the near future. Besides structured interviews and evaluations and test of technical skills, the motivation and understanding of a surgical career should be of additional focus. Both life-style factors and work-life balance impact attrition [44,163,164]. The fact that potential surgeons now will get less insight into surgical life, providing an early understanding of expectations not only for initial development, but also for a future surgical life, is important.

Evaluation of performance

One of the most challenging things regarding recruitment is to find measurements to evaluate performance and a positive outcome, since what is most important might differ from person to person. For the recruiter it might be to find someone who finishes residency without interruption and then continue to work as a specialist. For a supervisor it might be technical skills that make the resident independent at an early stage. For the surrounding team it might be the social skills, and for the patient a low rate of adverse events. All these factors are measurable, but the key question is which aspects are most important and at what stage, for instance, a lack of independence should be addressed. The question is obviously multifaceted. In other countries measures such as attrition and remediation have been used as predictors of performance. The current Swedish system does not allow for information of the current status of residents, of current recruitment processes, attrition, or remediation. Without standardization and continuous evaluations, it is not possible to know how the system performs or whether similar problems compared to other countries exist here as well. Sweden might have the perfect selection system - or

might have severe deficiencies. Besides actual numeric information on attrition the reason for it must be further investigated in a Swedish context. Given the prerequisites and working-climate at different departments, interviews with those leaving a residency or locum position should be done at every hospital since the challenges might differ. It might be related to working-climate or lack of support, or it might be related to a family situation that could not have been foreseen. In Ireland, a recent publication shows correlations between residents' performance at recruitment and progression to Specialty Training [142]. With the use of two selections, Ireland has an obvious and clear measure of success with the progression to Specialty Training. This reflects a successful recruitment, and not specifically higher competency compared to a Swedish doctor. Attrition rates and reasons for leaving, remediations, and time to becoming certified specialist would be simple markers of a successful recruitment, but it also assumes well-functioning evaluations during residency making sure the competences needed are fulfilled.

Diversity

Although striving for fair recruitment and equitable treatment, to the author´s knowledge, no studies on diversity in surgery have been conducted in Sweden. However, in a meta-analysis on callbacks (being offered an interview or a request for more information) including several professions, being a nonwhite applicant has been found to be a significant disadvantage in Sweden. Sweden was found to have the second highest level of discrimination of the nine countries included [165]. In the studies included in this thesis, 40 % of the included participants were women and race or ethnicity was not investigated. However, in the United States, the lack of diversity has been a recognized issue, though it has improved over the last few years [166]. Selection is a way to start improving this issue [167]. Incorporating the aspect of diversity could have been an improvement in the longitudinal ongoing study this thesis is part of, and it remains a promising area for future research, both in the recruitment aspect and in the surgical environment overall. From a recruitment perspective to ensure anonymity and decrease the likelihood of discrimination, applications could for example be conducted through systems enabling anonymity before decisions on who to select for an interview.

Artificial intelligence

In the last years, there has been an increasing interest in the use of Artificial Intelligence (AI) in surgery, with wide potential for improvement. Besides operative outcomes, AI could potentially be used to improve surgical education and recruitment [168,169]. Possible usages include incorporation into surgical simulators and provide real-time feedback, assist in identifying procedural steps or unsafe zones, and identify bias in the recruitment process [169]. While being in the early stage

of research, the field is growing and will likely expand in the coming years. In the Swedish setting with the decentralized recruitment and low number of applicants for each position the potential for assistance at point of recruitment seems limited. However, given the locum positions and lacking evaluation during these, the use of AI during this period, and during residency overall, forms a potential huge improvement. With increased feedback during surgery and repeated and gathered evaluations, this could potentially be transformative in surgical education. However, how and to which extent AI could be implemented while still ensuring safety is an area requiring additional research.

Future fair recruitment in residency

The interest in the recruitment process has increased in Sweden in recent years, as reflected by changes to the recruitment of interns [30]. While the exact method for evaluating future performance remains unknown, a standardized process would make the recruitment fairer and more transparent. Although there are differences between the recruitment of interns and residents - such as the larger number of positions to fill among interns - improvements could still be made in a similar way in the recruitment of residents. Based on the current research, there are simple steps that can be taken to improve the process. Firstly, the upcoming position should be advertised externally to ensure multiple candidates and candidates from outside the clinic. Secondly, decisions on who to call for an interview should be based on anonymized applications. Thirdly, use a structured and reproducible recruitment method that includes a structured interview. Finally, use a minimum of two assessors at the interview to reduce the risk of bias (Figure 8).

Figure 8. Fair future for residents. Suggestions to improve the transparency in the recruitment process for surgical residents.

Although there is no perfect selection method that can be concluded from the current data in this thesis, the longitudinal study that this thesis forms part of is ongoing and has the potential to provide increased clarity in the coming years. Increased transparency could be the first step in improving the current selection of surgical residents.

Strengths & limitations

The included studies have several strengths. This thesis is part of a longitudinal study, with a multidisciplinary research group with different specialities and expert areas, leading to the possibility for discussions and developments over several fields. For the specific papers there are additional strengths. In **Paper I** the used research triangulation and the collection of data over a longer time are beneficial. In **Paper II-IV** participants from seven hospitals were included, the rather long inclusion time, and that the tests are conducted separately from a recruitment process.

There are also several limitations in the conducted studies. In **Paper I**, there were limitations both in terms of language and number of responding UEMS board members, which affected the inclusion and differences in the thoroughness of the description of the countries´ selection processes. However, this study was carried out with the aim of describing the current selection processes and despite differences in detail, we obtained descriptions from almost all the countries in the European Union and were able to identify both similarities and dissimilarities.

Paper II-IV share several limitations. First, the data were collected outside of an actual selection process, which may have influenced their performance and responses. Knowing that this had no impact on their future employment, it could be that the participants were not fully investing in the tasks and therefore underperformed, simply forgot the tasks, or gave weak answers that they would not have given in a sharp employment situation. On the other hand, it could also be a strength, as it could make the answers in the interviews more honest and thus provide more truthful descriptions. Second, the possible selection bias due to that some applicants who did not receive a job declined to participate. Third, a more evenly distributed group between general surgery and urology would have been an improvement.

Paper II-III had some common limitations due to a rather small number of participants, which increases the risk of not being able to identify a true difference. In addition, **Paper II** lacked an objective measure of 3D cinema vision, and no alignment of raters was conducted with test videos in advance of their judgement. The latter will be a simple improvement to make in the future assessments in the longitudinal study.

In **Paper III** a limitation was that twelve participants did not receive any 360-degree evaluations, which may have affected the results. There are several possible explanations for this. First, the participants were unaware of how many had completed the assessments due to the system for distributing the survey based on codes. Second, not all had started work at the time of enrolment, and third, some had been working for a limited time. All these factors may have affected their willingness or ability to find assessors.

Acknowledgements

This thesis would not have been possible without the support of many people who all deserve a huge thank you.

First of all, my entire research group: you have taught me so much in how research works, on how to collaborate, how to think, and to see things from several perspectives. I have learned from all of you in different ways and are thankful for all the discussions and constructive feedback I have received over the years.

My (nowadays) main supervisor, Kristine Hagelsteen. You brought me into this project and eventually made me a part of it. I am grateful for all the opportunities I have had to learn, for conversations, but also for the kindness you have shown me over the years, not only in questions related to work.

My other two main supervisors during the study period, Anders Bergenfelz and Mikael Ekelund. All though maybe not always choosing to have the additional burden of a PhD student you have both shown me nothing but patients and support when needed.

Martin Bäckström for increasing my interest and curiosity around psychology, and for making statistics (relatively) easy to understand. Even though it sometimes took repeated lessons.

Chris Mathieu and Britt-Marie Johansson for guiding me through the (extremely unknown and diffuse) world of qualitative research. Your research world may be different from mine, but your knowledge and perspectives have broadened and increased my respect for the qualitative side.

Magnus Anderberg for teamwork, discussions, and improvements over the years.

Alexander Tejera for being additional support in the qualitative world, for planning and co-operation despite unmatched schedules.

All my colleagues from my clinical work who have listened, commented, enabled research time, and supported me through the years.

Det hade inte heller gått utan hjälp hemifrån, och jag är tacksam för alla familjemedlemmar och vänner som i perioder stått ut med mitt klagande. Tack Christoffer för att du har varit min balans i livet. För att du fullt ut förstår vikten av att fira, och för att du också förstår när det inte är läge att göra det. För att du gör kartor åt mig eller löser tekniska problem när mitt eget främsta alternativ är att

slänga datorn i väggen och aldrig mer öppna den. Det finns inte en chans att jag hade tagit mig igenom detta utan dig och jag hade inte heller velat det.

Isabelle och Jasmine, ni ger mig en värld utanför detta som är viktigare än allt. Och tack för att framsidan till den här boken är snyggare än om jag hade gjort den.

THANK YOU!

Errata

Paper I

Results – Under "Mapping". Section three.

Now reads: "Twelve countries have decentralized selection processes with recruitment, evaluation and selection decision-making by local universities and medical schools or hospitals."

Should be: "Eleven countries have decentralized selection processes with recruitment, evaluation and selection decision-making by local universities and medical schools or hospitals."

Paper II

Results - Table 4, for continuous outcome.

Now reads: "OR (95 % CI)"

Should be: Beta-coefficient (95 % CI)"

Paper III

Results – Under "Controlling for strong predictors"

Now reads: "There were four significant correlations between 360-degree ratings and the independent variables (Table 3). All significant correlations were to extraversion."

Should be: "There were six significant correlations between 360-degree ratings and the independent variables (Table 3). Four of the six were to extraversion."

References

1. Stulberg JJ, Huang R, Kreutzer L, et al. Association Between Surgeon Technical Skills and Patient Outcomes. *JAMA Surg*. Aug 19 2020;155(10):960-8. doi:10.1001/jamasurg.2020.3007
2. Mazzocco K, Petitti DB, Fong KT, et al. Surgical team behaviors and patient outcomes. *Am J Surg*. May 2009;197(5):678-85. doi:10.1016/j.amjsurg.2008.03.002
3. Skador i somatisk vård (Sveriges Kommuner och Regioner) (2024).
4. Gardner AK, Grantcharov T, Dunkin BJ. The Science of Selection: Using Best Practices From Industry to Improve Success in Surgery Training. *J Surg Educ*. Mar-Apr 2018;75(2):278-285. doi:10.1016/j.jsurg.2017.07.010
5. Specialistutbildningsfrågor SAfP-o. Läkares vidareutbildning i de nordiska länderna. 2010;
6. Organization WH. *Global patient safety action plan 2021–2030: towards eliminating avoidable harm in health care*. 2021.
7. Allvarliga skador och vårdskador (Socialstyrelsen) (2019).
8. Saka N, Yamamoto N, Watanabe J, et al. Comparison of Postoperative outcomes Among Patients Treated by Male Versus Female Surgeons: A Systematic Review and Meta-analysis. *Ann Surg*. May 10 2024;doi:10.1097/sla.0000000000006339
9. Blohm M, Sandblom G, Enochsson L, Österberg J. Differences in Cholecystectomy Outcomes and Operating Time Between Male and Female Surgeons in Sweden. *JAMA Surg*. Nov 1 2023;158(11):1168-1175. doi:10.1001/jamasurg.2023.3736
10. Engdahl J, Öberg A, Bech-Larsen S, et al. Short- and long-term outcome after colon cancer resections performed by male and female surgeons: A single-center retrospective cohort study. *Scand J Surg*. Feb 18 2024:14574969241228510. doi:10.1177/14574969241228510
11. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. Oct 10 2013;369(15):1434-42. doi:10.1056/NEJMsa1300625
12. Nyberg M, Sjoberg DD, Carlsson SV, et al. Surgeon heterogeneity significantly affects functional and oncological outcomes after radical prostatectomy in the Swedish LAPPRO trial. *BJU Int*. Mar 2021;127(3):361-368. doi:10.1111/bju.15238
13. Van den Broeck T, Oprea-Lager D, Moris L, et al. A Systematic Review of the Impact of Surgeon and Hospital Caseload Volume on Oncological and Nononcological Outcomes After Radical Prostatectomy for Nonmetastatic Prostate Cancer. *Eur Urol*. Nov 2021;80(5):531-545. doi:10.1016/j.eururo.2021.04.028

14. Cooper WO, Spain DA, Guillamondegui O, et al. Association of Coworker Reports About Unprofessional Behavior by Surgeons With Surgical Complications in Their Patients. *JAMA Surg*. Sep 1 2019;154(9):828-834. doi:10.1001/jamasurg.2019.1738
15. Christian CK, Gustafson ML, Roth EM, et al. A prospective study of patient safety in the operating room. *Surgery*. Feb 2006;139(2):159-73. doi:10.1016/j.surg.2005.07.037
16. Gjeraa K, Spanager L, Konge L, Petersen RH, Østergaard D. Non-technical skills in minimally invasive surgery teams: a systematic review. *Surg Endosc*. Dec 2016;30(12):5185-5199. doi:10.1007/s00464-016-4890-1
17. Woods MS, Liberman JN, Rui P, et al. Association between Surgical Technical Skills and Clinical Outcomes: A Systematic Literature Review and Meta-Analysis. *Jsls*. Jan-Mar 2023;27(1)doi:10.4293/jsls.2022.00076
18. Gardner AK, Smink DS, Scott BG, Korndorffer JR, Jr., Harrington D, Ritter EM. How Much Are We Spending on Resident Selection? *J Surg Educ*. Nov 2018;75(6):e85-e90. doi:10.1016/j.jsurg.2018.10.001
19. Khoushhal Z, Hussain MA, Greco E, et al. Prevalence and Causes of Attrition Among Surgical Residents: A Systematic Review and Meta-analysis. *JAMA Surg*. Mar 1 2017;152(3):265-272. doi:10.1001/jamasurg.2016.4086
20. Sundel M, Cairns C, Kavic SM. Cutting Ties: General Surgery Residents Have Higher Attrition Rates Than Residents in Other Surgical Training Programs. *J Surg Educ*. Jul 2024;81(7):900-904. doi:10.1016/j.jsurg.2024.04.004
21. Shweikeh F, Schwed AC, Hsu CH, Nfonsam VN. Status of Resident Attrition From Surgical Residency in the Past, Present, and Future Outlook. *J Surg Educ*. Mar-Apr 2018;75(2):254-262. doi:10.1016/j.jsurg.2017.07.015
22. Yaghoubian A, Galante J, Kaji A, et al. General surgery resident remediation and attrition: a multi-institutional study. *Arch Surg*. Sep 2012;147(9):829-33. doi:10.1001/archsurg.2012.1676
23. Schmidt FL, Hunter JE. The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*. Sep 1998;124(2):262-274. doi:10.1037/0033-2909.124.2.262
24. Highhouse S. Stubborn Reliance on Intuition and Subjectivity in Employee Selection. *Industrial and Organizational Psychology-Perspectives on Science and Practice*. Sep 2008;1(3):333-342. doi:10.1111/j.1754-9434.2008.00058.x
25. *Pilot's Handbook of Aeronautical Knowledge FAA-H-8083-25C*. United States Department of Transportation, Federal Aviation Administration, Airman, Testing Standards Branch; 2023. https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/faa-h-8083-25c.pdf
26. Paice AG, Aggarwal R, Darzi A. Safety in surgery: is selection the missing link? *World J Surg*. Sep 2010;34(9):1993-2000. doi:10.1007/s00268-010-0619-9
27. Core Surgical Training (CST). RCSI Royal College of Surgeons in Ireland. Accessed Feb 11, 2025. https://www.rcsi.com/surgery/training/surgery/cst/apply
28. SPECIALTY TRAINING INTAKE GUIDE 2025 APPLICANT EDITION

29. Hagelsteen K, Johansson BM, Bergenfelz A, Mathieu C. Identification of Warning Signs During Selection of Surgical Trainees. *J Surg Educ*. May-Jun 2019;76(3):684-693. doi:10.1016/j.jsurg.2018.12.002
30. Sveriges Yngre Läkares Förening S. Schysst Rekrytering av AT-läkare. Sveriges läkarförbund. Accessed Feb 26, 2025. https://slf.se/sylf/din-karriar/at-lakare/schysst-rekrytering/
31. Makdisi G, Takeuchi T, Rodriguez J, Rucinski J, Wise L. How we select our residents--a survey of selection criteria in general surgery residents. *J Surg Educ*. Jan-Feb 2011;68(1):67-72. doi:10.1016/j.jsurg.2010.10.003
32. Marwan Y, Waly F, Algarni N, Addar A, Saran N, Snell L. The Role of Letters of Recommendation in the Selection Process of Surgical Residents in Canada: A National Survey of Program Directors. *J Surg Educ*. Jul-Aug 2017;74(4):762-767. doi:10.1016/j.jsurg.2017.01.006
33. Stephenson-Famy A, Houmard BS, Oberoi S, Manyak A, Chiang S, Kim S. Use of the Interview in Resident Candidate Selection: A Review of the Literature. *J Grad Med Educ*. Dec 2015;7(4):539-48. doi:10.4300/jgme-d-14-00236.1
34. Wiesner WH, Cronshaw SF. A META-ANALYTIC INVESTIGATION OF THE IMPACT OF INTERVIEW FORMAT AND DEGREE OF STRUCTURE ON THE VALIDITY OF THE EMPLOYMENT INTERVIEW. *Journal of Occupational Psychology*. Dec 1988;61(4):275-290. doi:10.1111/j.2044-8325.1988.tb00467.x
35. Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ*. Mar 2004;38(3):314-26. doi:10.1046/j.1365-2923.2004.01776.x
36. Roberts C, Clark T, Burgess A, Frommer M, Grant M, Mossman K. The validity of a behavioural multiple-mini-interview within an assessment centre for selection into specialty training. *BMC Med Educ*. Aug 13 2014;14:169. doi:10.1186/1472-6920-14-169
37. Dore KL, Kreuger S, Ladhani M, et al. The reliability and acceptability of the Multiple Mini-Interview as a selection instrument for postgraduate admissions. *Acad Med*. Oct 2010;85(10 Suppl):S60-3. doi:10.1097/ACM.0b013e3181ed442b
38. Lin JC, Hu DJ, Scott IU, Greenberg PB. Evidence-Based Practices for Interviewing Graduate Medical Education Applicants: A Systematic Review. *J Grad Med Educ*. Apr 2024;16(2):151-165. doi:10.4300/jgme-d-23-00115.1
39. Patterson F, Zibarras L, Ashworth V. Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. *Med Teach*. 2016;38(1):3-17. doi:10.3109/0142159x.2015.1072619
40. Webster ES, Paton LW, Crampton PES, Tiffin PA. Situational judgement test validity for selection: A systematic review and meta-analysis. *Med Educ*. Oct 2020;54(10):888-902. doi:10.1111/medu.14201
41. 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*. Jan 2018;40(1):3-19. doi:10.1080/0142159x.2017.1367375

42. Schaverien MV. Selection for Surgical Training: An Evidence-Based Review. *J Surg Educ*. Jul-Aug 2016;73(4):721-9. doi:10.1016/j.jsurg.2016.02.007
43. Louridas M, Szasz P, de Montbrun S, Harris KA, Grantcharov TP. Can We Predict Technical Aptitude?: A Systematic Review. *Ann Surg*. Apr 2016;263(4):673-91. doi:10.1097/sla.0000000000001283
44. Alterman DM, Jones TM, Heidel RE, Daley BJ, Goldman MH. The predictive value of general surgery application data for future resident performance. *J Surg Educ*. Nov-Dec 2011;68(6):513-8. doi:10.1016/j.jsurg.2011.07.007
45. Labiner HE, Anderson CE, Maloney Patel N. Virtual Recruitment in Surgical Residency Programs. *Curr Surg Rep*. 2021;9(11):25. doi:10.1007/s40137-021-00302-9
46. Daniel M, Gottlieb M, Wooten D, et al. Virtual interviewing for graduate medical education recruitment and selection: A BEME systematic review: BEME Guide No. 80. *Medical Teacher*. Dec 2022;44(12):1313-1331. doi:10.1080/0142159x.2022.2130038
47. Wright AS. Virtual Interviews for Fellowship and Residency Applications Are Effective Replacements for In-Person Interviews and Should Continue Post-COVID. *J Am Coll Surg*. Dec 2020;231(6):678-680. doi:10.1016/j.jamcollsurg.2020.09.005
48. Louridas M, Szasz P, Fecso AB, et al. Practice does not always make perfect: need for selection curricula in modern surgical training. *Surg Endosc*. Sep 2017;31(9):3718-3727. doi:10.1007/s00464-017-5572-3
49. Drebin HM, Cohen ZC, Anderson CB, Badalato GM. Innate Versus Acquired: A Review of Predictive Technical Aptitude Assessments in Surgical Trainee Selection. *Curr Urol Rep*. Jan 14 2025;26(1):25. doi:10.1007/s11934-024-01253-1
50. Grantcharov TP, Funch-Jensen P. Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. *Am J Surg*. Apr 2009;197(4):447-9. doi:10.1016/j.amjsurg.2008.01.024
51. Buckley CE, Kavanagh DO, Nugent E, Ryan D, Traynor OJ, Neary PC. The impact of aptitude on the learning curve for laparoscopic suturing. *Am J Surg*. Feb 2014;207(2):263-70. doi:10.1016/j.amjsurg.2013.08.037
52. Buckley CE, Kavanagh DO, Gallagher TK, Conroy RM, Traynor OJ, Neary PC. Does aptitude influence the rate at which proficiency is achieved for laparoscopic appendectomy? *J Am Coll Surg*. Dec 2013;217(6):1020-7. doi:10.1016/j.jamcollsurg.2013.07.405
53. Bosker R, Groen H, Hoff C, Totte E, Ploeg R, Pierie JP. Early learning effect of residents for laparoscopic sigmoid resection. *J Surg Educ*. Mar-Apr 2013;70(2):200-5. doi:10.1016/j.jsurg.2012.10.004
54. 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*. Jun 2007;193(6):797-804. doi:10.1016/j.amjsurg.2006.06.050
55. Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ. A systematic review of skills transfer after surgical simulation training. *Ann Surg*. Aug 2008;248(2):166-79. doi:10.1097/SLA.0b013e318176bf24

56. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg*. Oct 2002;236(4):458-63; discussion 463-4. doi:10.1097/00000658-200210000-00008
57. Thomaschewski M, Vonthein R, Keck T, Laubert T, Benecke C. Laparoscopic simulation training improves operating room performance of surgical residents: a multicenter randomized trial (NOVICE). *Int J Surg*. Feb 24 2025;doi:10.1097/js9.0000000000002304
58. Munz Y, Kumar BD, Moorthy K, Bann S, Darzi A. Laparoscopic virtual reality and box trainers: is one superior to the other? *Surg Endosc*. Mar 2004;18(3):485-94. doi:10.1007/s00464-003-9043-7
59. Zwimpfer TA, Lacher D, Fellmann-Fischer B, Mueller M. A laparoscopic study investigating 3D vs 2D imaging systems using a pelvitrainer model with experts, non-experts, and students. *BMC Surg*. Nov 9 2020;20(1):276. doi:10.1186/s12893-020-00892-8
60. Hagelsteen K, Langegård A, Lantz A, Ekelund M, Anderberg M, Bergenfelz A. Faster acquisition of laparoscopic skills in virtual reality with haptic feedback and 3D vision. *Minim Invasive Ther Allied Technol*. Oct 2017;26(5):269-277. doi:10.1080/13645706.2017.1305970
61. Österberg J, Sokratous A, Georgiou K, Enochsson L. Can 3D Vision Improve Laparoscopic Performance in Box Simulation Training when Compared to Conventional 2D Vision? *Folia Med (Plovdiv)*. Dec 31 2019;61(4):491-499. doi:10.3897/folmed.61.e47958
62. Restaino S, Scutiero G, Taliento C, et al. Three-dimensional vision versus two-dimensional vision on laparoscopic performance of trainee surgeons: a systematic review and meta-analysis. *Updates Surg*. Apr 2023;75(3):455-470. doi:10.1007/s13304-023-01465-z
63. El Boghdady M, Ewalds-Kvist BM. The innate aptitude's effect on the surgical task performance: a systematic review. *Updates Surg*. Sep 25 2021;doi:10.1007/s13304-021-01173-6
64. Maan ZN, Maan IN, Darzi AW, Aggarwal R. Systematic review of predictors of surgical performance. *Br J Surg*. Dec 2012;99(12):1610-21. doi:10.1002/bjs.8893
65. Mason EM, Deal MJ, Richey BP, et al. Innate Arthroscopic & Laparoscopic Surgical Skills: A Systematic Review of Predictive Performance Indicators Within Novice Surgical Trainees. *J Surg Educ*. Jan-Feb 2021;78(1):178-200. doi:10.1016/j.jsurg.2020.06.006
66. Beattie KL, Hill A, Horswill MS, Grove PM, Stevenson ARL. Aptitude and attitude: predictors of performance during and after basic laparoscopic skills training. *Surg Endosc*. Aug 9 2021:1-13. doi:10.1007/s00464-021-08668-8
67. Brandt MG, Davies ET. Visual-spatial ability, learning modality and surgical knot tying. *Can J Surg*. Dec 2006;49(6):412-6.

68. Gallagher AG, Cowie R, Crothers I, Jordan-Black JA, Satava RM. PicSOr: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surg Endosc*. Sep 2003;17(9):1468-71. doi:10.1007/s00464-002-8569-4
69. Vajsbaher T, Schultheis H, Janssen S, et al. The development of visuospatial abilities and their impact on laparoscopic skill acquisition: a clinical longitudinal study. *Surg Endosc*. May 31 2022:1-10. doi:10.1007/s00464-022-09328-1
70. Maurice-Ventouris M, Muller Moran HR, Alharbi M, Ahn BT, Harley JM, Lachapelle KJ. The study of visuospatial abilities in trainees: A scoping review and proposed model. *Surg Open Sci*. Jul 2021;5:25-33. doi:10.1016/j.sopen.2021.05.001
71. Kalun P, Dunn K, Wagner N, Pulakunta T, Sonnadara R. Recent evidence on visual-spatial ability in surgical education: A scoping review. *Can Med Educ J*. Dec 2020;11(6):e111-e127. doi:10.36834/cmej.69051
72. Stefanidis D, Korndorffer JR, Jr., Black FW, et al. Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. *Surgery*. Aug 2006;140(2):252-62. doi:10.1016/j.surg.2006.04.002
73. Causby R, Reed L, McDonnell M, Hillier S. Use of objective psychomotor tests in health professionals. *Percept Mot Skills*. Jun 2014;118(3):765-804. doi:10.2466/25.27.PMS.118k27w2
74. Van Herzeele I, O'Donoghue KG, Aggarwal R, Vermassen F, Darzi A, Cheshire NJ. Visuospatial and psychomotor aptitude predicts endovascular performance of inexperienced individuals on a virtual reality simulator. *J Vasc Surg*. Apr 2010;51(4):1035-42. doi:10.1016/j.jvs.2009.11.059
75. Louridas M, Quinn LE, Grantcharov TP. Predictive value of background experiences and visual spatial ability testing on laparoscopic baseline performance among residents entering postgraduate surgical training. *Surg Endosc*. Mar 2016;30(3):1126-33. doi:10.1007/s00464-015-4313-8
76. Lin D, Pena G, Field J, et al. What are the demographic predictors in laparoscopic simulator performance? *ANZ J Surg*. Dec 2016;86(12):983-989. doi:10.1111/ans.12992
77. Prineas S, Mosier K, Mirko C, Guicciardi S. Non-technical Skills in Healthcare. In: Donaldson L, Ricciardi W, Sheridan S, Tartaglia R, eds. *Textbook of Patient Safety and Clinical Risk Management*. Springer. Copyright 2021, The Author(s). 2021:413-34.
78. Bruinsma WE, Becker SJ, Guitton TG, Kadzielski J, Ring D. How prevalent are hazardous attitudes among orthopaedic surgeons? *Clin Orthop Relat Res*. May 2015;473(5):1582-9. doi:10.1007/s11999-014-3966-2
79. Kadzielski J, McCormick F, Herndon JH, Rubash H, Ring D. Surgeons' attitudes are associated with reoperation and readmission rates. *Clin Orthop Relat Res*. May 2015;473(5):1544-51. doi:10.1007/s11999-014-3687-6
80. Sharma B, Mishra A, Aggarwal R, Grantcharov TP. Non-technical skills assessment in surgery. *Surg Oncol*. Sep 2011;20(3):169-77. doi:10.1016/j.suronc.2010.10.001

81. Pradarelli JC, Gupta A, Hermosura AH, et al. Non-technical skill assessments across levels of US surgical training. *Surgery*. Sep 2021;170(3):713-718. doi:10.1016/j.surg.2021.02.058
82. Hughes BD, Perone JA, Cummins CB, et al. Personality Testing May Identify Applicants Who Will Become Successful in General Surgery Residency. *J Surg Res*. Jan 2019;233:240-248. doi:10.1016/j.jss.2018.08.003
83. Rothstein MG, Goffin RD. The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*. 2006;16(2):155-180. doi:10.1016/j.hrmr.2006.03.004
84. Roberts BW, DelVecchio WF. The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychol Bull*. Jan 2000;126(1):3-25. doi:10.1037/0033-2909.126.1.3
85. Allemand M, Gomez V, Jackson JJ. Personality trait development in midlife: exploring the impact of psychological turning points. *Eur J Ageing*. Sep 2010;7(3):147-155. doi:10.1007/s10433-010-0158-0
86. McCrae RR, John OP. An introduction to the five-factor model and its applications. *J Pers*. Jun 1992;60(2):175-215. doi:10.1111/j.1467-6494.1992.tb00970.x
87. Hoffman BM, Coons MJ, Kuo PC. Personality differences between surgery residents, nonsurgery residents, and medical students. *Surgery*. Aug 2010;148(2):187-93. doi:10.1016/j.surg.2010.04.005
88. Drosdeck JM, Osayi SN, Peterson LA, Yu L, Ellison EC, Muscarella P. Surgeon and nonsurgeon personalities at different career points. *J Surg Res*. Jun 1 2015;196(1):60-6. doi:10.1016/j.jss.2015.02.021
89. Bexelius TS, Olsson C, Järnbert-Pettersson H, Parmskog M, Ponzer S, Dahlin M. Association between personality traits and future choice of specialisation among Swedish doctors: a cross-sectional study. *Postgrad Med J*. Aug 2016;92(1090):441-6. doi:10.1136/postgradmedj-2015-133478
90. Bisset CN, McKee T, Cawley M, Tilling E, Moug SJ. A Systematic Review of the Abdominal Surgeon's Personality: Exploring Common Traits in Western Populations. *Behav Sci (Basel)*. Dec 26 2020;11(1)doi:10.3390/bs11010002
91. Sier VQ, Schmitz RF, Schepers A, van der Vorst JR. Exploring the surgical personality. *Surgeon*. Feb 2023;21(1):1-7. doi:10.1016/j.surge.2022.01.008
92. Sier VQ, Schmitz RF, Putter H, Schepers A, van der Vorst JR. The big five: Studying the surgical personality. *Surgery*. Nov 2022;172(5):1358-1363. doi:10.1016/j.surg.2022.08.003
93. Whitaker M. The surgical personality: does it exist? *Ann R Coll Surg Engl*. Jan 2018;100(1):72-77. doi:10.1308/rcsann.2017.0200
94. Sier VQ, Bisset CN, Tesselaar DAJ, et al. Dissecting the surgeon's personality: cross-cultural comparisons in Western Europe. *Colorectal Dis*. Jun 2024;26(6):1239-1249. doi:10.1111/codi.16993
95. Sier VQ, Schmitz RF, Wertenbroek R, Schepers A, van der Vorst JR. Surgeon personality diversity across generations and subspecialties. *Surgeon*. Feb 17 2025;doi:10.1016/j.surge.2025.02.003

96. CA L, SAM N. Surgeons' personalities and surgical outcomes. *The Bulletin of the Royal College of Surgeons of England*. 2018;100(6):259-263. doi:10.1308/rcsbull.2018.259
97. Rosenthal R, Schäfer J, Hoffmann H, Vitz M, Oertli D, Hahnloser D. Personality traits and virtual reality performance. *Surg Endosc*. Jan 2013;27(1):222-30. doi:10.1007/s00464-012-2424-z
98. Teunis T, Janssen SJ, Guitton TG, Vranceanu AM, Goos B, Ring D. Surgeon personality is associated with recommendation for operative treatment. *Hand (N Y)*. Dec 2015;10(4):779-84. doi:10.1007/s11552-015-9755-x
99. Travers V. Burnout in orthopedic surgeons. *Orthop Traumatol Surg Res*. Feb 2020;106(1s):S7-s12. doi:10.1016/j.otsr.2019.04.029
100. Balch CM, Freischlag JA, Shanafelt TD. Stress and burnout among surgeons: understanding and managing the syndrome and avoiding the adverse consequences. *Arch Surg*. Apr 2009;144(4):371-6. doi:10.1001/archsurg.2008.575
101. Prins DJ, van Vendeloo SN, Brand PLP, et al. The relationship between burnout, personality traits, and medical specialty. A national study among Dutch residents. *Med Teach*. May 2019;41(5):584-590. doi:10.1080/0142159x.2018.1514459
102. Kalet A, Guerrasio J, Chou CL. Twelve tips for developing and maintaining a remediation program in medical education. *Med Teach*. Aug 2016;38(8):787-92. doi:10.3109/0142159x.2016.1150983
103. Hagelsteen K, Mathieu C. Resident selection in surgery. 2023;
104. Ajjawi R, Bearman M, Molloy E, Noble C. The role of feedback in supporting trainees who underperform in clinical environments. *Front Med (Lausanne)*. 2023;10:1121602. doi:10.3389/fmed.2023.1121602
105. El Boghdady M, Alijani A. Feedback in surgical education. *Surgeon*. Apr 2017;15(2):98-103. doi:10.1016/j.surge.2016.06.006
106. Cantillon P, Sargeant J. Giving feedback in clinical settings. *Bmj*. Nov 10 2008;337:a1961. doi:10.1136/bmj.a1961
107. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. Article. *Medical teacher*. 2010;32(8):638-645.
108. Iobst WF, Sherbino J, Cate OT, et al. Competency-based medical education in postgraduate medical education. *Med Teach*. 2010;32(8):651-6. doi:10.3109/0142159x.2010.500709
109. Jug R, Jiang XS, Bean SM. Giving and Receiving Effective Feedback: A Review Article and How-To Guide. *Arch Pathol Lab Med*. Feb 2019;143(2):244-250. doi:10.5858/arpa.2018-0058-RA
110. Utbildningskommittén SKF. Utbildningsbok Kirurgi. Accessed Jul 18, 2025. https://svenskkirurgiskforening.se/wp-content/uploads/2023/04/Utbildningsboken_2022.pdf
111. Ott MC, Pack R, Cristancho S, Chin M, Van Koughnett JA, Ott M. "The Most Crushing Thing": Understanding Resident Assessment Burden in a Competency-Based Curriculum. *Journal of Graduate Medical Education*. 2022;14(5):583-592. doi:10.4300/jgme-d-22-00050.1

112. Szulewski A, Braund H, Dagnone DJ, et al. The Assessment Burden in Competency-Based Medical Education: How Programs Are Adapting. *Academic Medicine*. 2023;98(11):1261-1267. doi:10.1097/acm.0000000000005305
113. Lägervik M, Thörne K, Fristedt S, Henricson M, Hedberg B. Residents' and supervisors' experiences when using a feedback-model in post-graduate medical education. *BMC Med Educ*. Dec 23 2022;22(1):891. doi:10.1186/s12909-022-03969-5
114. Shafian S, Ilaghi M, Shahsavani Y, et al. The feedback dilemma in medical education: insights from medical residents' perspectives. *BMC Med Educ*. Apr 19 2024;24(1):424. doi:10.1186/s12909-024-05398-y
115. Sender Liberman A, Liberman M, Steinert Y, McLeod P, Meterissian S. Surgery residents and attending surgeons have different perceptions of feedback. *Med Teach*. Aug 2005;27(5):470-2. doi:10.1080/0142590500129183
116. Yarris LM, Linden JA, Gene Hern H, et al. Attending and resident satisfaction with feedback in the emergency department. *Acad Emerg Med*. Dec 2009;16 Suppl 2:S76-81. doi:10.1111/j.1553-2712.2009.00592.x
117. Raikhel AV, Starks H, Berger G, Redinger J. Through the Looking Glass: Comparing Hospitalists' and Internal Medicine Residents' Perceptions of Feedback. *Cureus*. Jun 2024;16(6):e63459. doi:10.7759/cureus.63459
118. Hattie J, Timperley H. The power of feedback. *Review of Educational Research*. Mar 2007;77(1):81-112. doi:10.3102/003465430298487
119. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. Apr 2009;42(2):377-81. doi:10.1016/j.jbi.2008.08.010
120. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform*. Jul 2019;95:103208. doi:10.1016/j.jbi.2019.103208
121. Manual for Kit of Factor-Referenced Cognitive Tests Educational Testing Service; 1976.
122. Purdue Pegboard Test. Lafayette Instrument Company; 2015.
123. LAPSIM® – YOUR FIRST CHOICE FOR LAPAROSCOPY TRAINING. Surgical Science. Accessed 10 January, 2022. https://surgicalscience.com/simulators/lapsim/
124. SIMBALL BOX. Surgical Science. Accessed 10 January, 2022. https://surgicalscience.com/simulators/simball-box/
125. Moorthy K, Munz Y, Dosis A, Bello F, Chang A, Darzi A. Bimodal assessment of laparoscopic suturing skills: construct and concurrent validity. *Surg Endosc*. Nov 2004;18(11):1608-12. doi:10.1007/s00464-003-9312-5
126. Bäckström M, Björklund F, Maddux RE, Lindén M. The NB5I: A full-scale Big-Five inventory with evaluatively neutralized items. *European Journal of Psychological Assessment*. 2023:No Pagination Specified-No Pagination Specified. doi:10.1027/1015-5759/a000687

127. Archer JC, Norcini J, Davies HA. Use of SPRAT for peer review of paediatricians in training. *Bmj*. May 28 2005;330(7502):1251-3. doi:10.1136/bmj.38447.610451.8F
128. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006;3(2):77-101. doi:10.1191/1478088706qp063oa
129. Frambach JM, van der Vleuten CPM, Durning SJ. AM Last Page: Quality Criteria in Qualitative and Quantitative Research. *Academic Medicine*. 2013;88(4):552. doi:10.1097/ACM.0b013e31828abf7f
130. Varpio L, Ajjawi R, Monrouxe LV, O'Brien BC, Rees CE. Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking. *Med Educ*. Jan 2017;51(1):40-50. doi:10.1111/medu.13124
131. Royal College of Surgeons in Ireland. General Surgery. Royal College of Surgeons in Ireland,. Jan 3, 2021. https://www.rcsi.com/surgery/training/surgery/general-surgery/overview
132. Linstead S, Fulop, L., Lilley, S.. *Management and Organization*. 2 ed. Palgrave Macmillan; 2009.
133. Selber JC, Tong W, Koshy J, Ibrahim A, Liu J, Butler C. Correlation between trainee candidate selection criteria and subsequent performance. *J Am Coll Surg*. Nov 2014;219(5):951-7. doi:10.1016/j.jamcollsurg.2014.07.942
134. Towaij C, Gawad N, Alibhai K, Doan D, Raîche I. Trust Me, I Know Them: Assessing Interpersonal Bias in Surgery Residency Interviews. *J Grad Med Educ*. Jun 2022;14(3):289-294. doi:10.4300/jgme-d-21-00882.1
135. Strand EA, Moore E, Laube DW. Can a structured, behavior-based interview predict future resident success? *Am J Obstet Gynecol*. May 2011;204(5):446.e1-446.e13. doi:10.1016/j.ajog.2011.02.019
136. Marcus-Blank B, Dahlke JA, Braman JP, et al. Predicting Performance of First-Year Residents: Correlations Between Structured Interview, Licensure Exam, and Competency Scores in a Multi-Institutional Study. *Acad Med*. Mar 2019;94(3):378-387. doi:10.1097/acm.0000000000002429
137. Rosenfeld JM, Reiter HI, Trinh K, Eva KW. A cost efficiency comparison between the multiple mini-interview and traditional admissions interviews. *Adv Health Sci Educ Theory Pract*. Mar 2008;13(1):43-58. doi:10.1007/s10459-006-9029-z
138. Gardner AK, Dunkin BJ. Pursuing Excellence: The Power of Selection Science to Provide Meaningful Data and Enhance Efficiency in Selecting Surgical Trainees. *Ann Surg*. Jul 2019;270(1):188-192. doi:10.1097/sla.0000000000002806
139. Bergen PC, Littlefield JH, O'Keefe GE, et al. Identification of high-risk residents. *J Surg Res*. Aug 2000;92(2):239-44. doi:10.1006/jsre.2000.5924
140. Williams RG, Roberts NK, Schwind CJ, Dunnington GL. The nature of general surgery resident performance problems. *Surgery*. Jun 2009;145(6):651-8. doi:10.1016/j.surg.2009.01.019
141. Resnick AS, Mullen JL, Kaiser LR, Morris JB. Patterns and predictions of resident misbehavior--a 10-year retrospective look. *Curr Surg*. Nov-Dec 2006;63(6):418-25. doi:10.1016/j.cursur.2006.05.004

142. Toale C, Morris M, Gross S, et al. Performance in Irish Selection and Future Performance in Surgical Training. *JAMA Surg*. May 1 2024;159(5):538-545. doi:10.1001/jamasurg.2024.0034
143. Guide to core surgical training intake 2025 (Royal College of Surgeons in Ireland)
144. Toale C, Morris M, Kavanagh D. Perceptions and experiences of simulation-based assessment of technical skill in surgery: A scoping review. *Am J Surg*. Oct 2021;222(4):723-730. doi:10.1016/j.amjsurg.2021.03.035
145. Selva Raj DR, Kumar S, Nallathamby K, Raj K, Hristova M. A Systematic Review of the Learning Curves of Novices and Trainees to Achieve Proficiency in Laparoscopic Skills: Virtual Reality Simulator Versus Box Trainer. *Cureus*. Nov 2024;16(11):e72923. doi:10.7759/cureus.72923
146. Hojat M, Erdmann JB, Gonnella JS. Personality assessments and outcomes in medical education and the practice of medicine: AMEE Guide No. 79. *Med Teach*. Jul 2013;35(7):e1267-301. doi:10.3109/0142159x.2013.785654
147. Thorndike EL. A constant error in psychological ratings. *Journal of Applied Psychology*. 1920;4(1):25-29. doi:10.1037/h0071663
148. Risucci DA, Lutsky L, Rosati RJ, Tortolani AJ. Reliability and accuracy of resident evaluations of surgical faculty. *Eval Health Prof*. Sep 1992;15(3):313-24. doi:10.1177/016327879201500304
149. Talamas SN, Mavor KI, Perrett DI. Blinded by Beauty: Attractiveness Bias and Accurate Perceptions of Academic Performance. *PLoS One*. 2016;11(2):e0148284. doi:10.1371/journal.pone.0148284
150. McKinstry BH, Cameron HS, Elton RA, Riley SC. Leniency and halo effects in marking undergraduate short research projects. *BMC Med Educ*. Nov 29 2004;4(1):28. doi:10.1186/1472-6920-4-28
151. Thomas MR, Beckman TJ, Mauck KF, Cha SS, Thomas KG. Group assessments of resident physicians improve reliability and decrease halo error. *J Gen Intern Med*. Jul 2011;26(7):759-64. doi:10.1007/s11606-011-1670-4
152. Hesketh EA, Anderson F, Bagnall GM, et al. Using a 360 degrees diagnostic screening tool to provide an evidence trail of junior doctor performance throughout their first postgraduate year. *Med Teach*. May 2005;27(3):219-33. doi:10.1080/01421590500098776
153. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach*. Nov 2007;29(9):855-71. doi:10.1080/01421590701775453
154. Henkelman E, Tornetta P, 3rd. Can Orthopaedic Residents Beat a Personality Assessment? *J Surg Educ*. Mar-Apr 2022;79(2):531-534. doi:10.1016/j.jsurg.2021.09.020
155. Bäckström M, Björklund F, Maddux RE, Lindén M. The NB5I: A full-scale Big-Five inventory with evaluatively neutralized items. *European Journal of Psychological Assessment*. 2023;39(2):132-140. doi:10.1027/1015-5759/a000687

156. Reddy ST, Zegarek MH, Fromme HB, Ryan MS, Schumann SA, Harris IB. Barriers and Facilitators to Effective Feedback: A Qualitative Analysis of Data From Multispecialty Resident Focus Groups. *J Grad Med Educ*. Jun 2015;7(2):214-9. doi:10.4300/jgme-d-14-00461.1
157. SYLF Sylf. När underläkarna tystnar. Accessed Nov 4, 2024. https://slf.se/sylf/app/uploads/2021/02/tyst-lakarkar-en-rapport-om-tystnadskultur-bland-underlakarna.pdf
158. Grundutbildning. Sveriges läkarförbund. Accessed Feb 10, 2025. https://slf.se/rad-och-stod/utbildning/grundutbildning/
159. Bastjänstgöring, BT. Sveriges läkarförbund. Jan 29, 2025. https://slf.se/rad-och-stod/utbildning/bastjanstgoring-bt/
160. Nya regler för läkarnas ST, inkl. BT. Socialstyrelsen. Accessed Feb 10, 2025. https://legitimation.socialstyrelsen.se/bastjanstgoring-bt/
161. Sveriges Yngre Läkares Förening S. Bastjänstgöring, BT. Sveriges Läkarförbund. Jan 29, 2025. https://slf.se/sylf/din-karriar/bastjanstgoring-bt/
162. Allmäntjänstgöring, AT. Sveriges Läkarförbund. Jan 29, 2025. https://slf.se/rad-och-stod/utbildning/allmantjanstgoring-at/
163. Hope C, Reilly JJ, Griffiths G, Lund J, Humes D. Factors Associated with Attrition and Performance Throughout Surgical Training: A Systematic Review and Meta-Analysis. *World J Surg*. Feb 2021;45(2):429-442. doi:10.1007/s00268-020-05844-0
164. Nally DM, Kavanagh D, Traynor O, Bennett D. Attrition From a National Surgical Training Program: A Qualitative Analysis. *Ann Surg*. Apr 1 2022;275(4):621-628. doi:10.1097/sla.0000000000004852
165. Hexel O, Quillian L, Heath A, Pager D, Midtbøen A, Fleischmann F. Do Some Countries Discriminate More than Others? Evidence from 97 Field Experiments of Racial Discrimination in Hiring. *Sociological Science*. 2019-06 2019;doi:10.15195/v6.a18
166. Oseni TO, Kelly BN, Pei K, et al. Diversity efforts in surgery: Are we there yet? *Am J Surg*. Jul 2022;224(1 Pt B):259-263. doi:10.1016/j.amjsurg.2022.01.014
167. Gardner AK. How Can Best Practices in Recruitment and Selection Improve Diversity in Surgery? *Ann Surg*. Jan 2018;267(1):e1-e2. doi:10.1097/sla.0000000000002496
168. Varghese C, Harrison EM, O'Grady G, Topol EJ. Artificial intelligence in surgery. *Nat Med*. May 2024;30(5):1257-1268. doi:10.1038/s41591-024-02970-3
169. Leon S, Lee S, Perez JE, Hashimoto DA. Artificial intelligence and the education of future surgeons. *Am J Surg*. Feb 19 2025:116257. doi:10.1016/j.amjsurg.2025.116257

Printed by Media-Tryck, Lund 2025 NORDIC SWAN ECOLABEL 3041 0903

Surgical Education

A doctor is made of several skills, or "pieces", and there is continuous development throughout their career. For surgeons, both technical and non-technical skills must be mastered. This thesis focuses on selection methods for surgical residents from both aspects and aims to contribute to the future development of the current selection process in Sweden.

Pediatrics
Department of Clinical Sciences, Lund

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2025:126
ISBN 978-91-8021-779-8
ISSN 1652-8220