Making the cut: An evaluation of selection into the Royal Australasian College of Surgeons' surgical training program

Zaita Oldfield

Student ID 3627973

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Education

College of Arts and Education

Victoria University

Melbourne, Australia

October 2018

Abstract

Selection of trainees for surgical training is widely acknowledged to be both complex and important. Doctors embarking on surgical careers expend considerable time and resources in training, as do those responsible for facilitating their learning. The aim of surgical training is to prepare surgeons to be competent, effective, ethical practitioners in unsupervised clinical practice as consultant surgeons (Carroll, Kennedy, Traynor & Gallagher, 2009; Elfenbein, Sippel, McDonald, Watson, Scarborough & Migaly, 2015). The aim of selection must be to admit those who are most likely to succeed in surgical training and beyond.

Admitting candidates who are inadequately suited to surgical training can result in trainees who unduly struggle or are unable to satisfy training requirements, and may ultimately jeopardise patient safety. In contrast, not accepting candidates who are well-suited to surgical training may be unfair to both candidates and communities that miss out on skillful surgeons. These aspects are among those that make selection to surgical training an extremely 'high-stakes' activity.

Those responsible for selection grapple with assessing candidates' current skills and attributes, and with ascertaining their likely future performance. Limitations of, and tensions between interacting elements—human agency, requirement specifications, instruments, processes, influences—in selection to surgical training, mean that the long-term outcomes of selection can never be assured. This study identifies principal elements in selection to surgical training, discusses influences on, and interrelationships between the elements, and reviews connections between selection and surgical training assessments.

This study has appraised the current instruments used for selection into the Royal Australasian College of Surgeons' (RACS) General Surgery (GS) training program in Australia and New Zealand, to establish their effectiveness in predicting trainees' performance in assessments during the first two years of training. Data were considered for selection and assessment items for trainees for three yearly cohorts, selected in 2008, 2009, and 2010. The

study compared the performance of trainees in three selection instruments—a structured CV, a structured referee report, and a multi-station interview—to their performance in three examinations and three work-based assessments.

Firstly, Pearson product-moment correlations were calculated to examine the extent of relationships within each set of variables to determine intra-relationships of the selection items, of the examination items, and of the work-based assessment items. Secondly, Pearson product-moment correlations were conducted to determine degrees of association between selection items and performance in each of the subsequent assessment items. Thirdly, multiple regression analysis was conducted to determine the extent to which trainees' scores in the selection items (independent variables) predicted scores in each of the assessments during training (dependent variables). The model fit and strength was assessed using the analysis of variance (ANOVA) step within the regression analysis. The relative strength of the associations between dependent variables and the independent variables were assessed using regression coefficients.

The findings of this study have shown that performance in RACS GS selection partially predicts performance in assessments during training. In general, candidates' performance varied across each of the selection items, performance in all examinations was highly consistent and performance in major end of term work-based assessments was also consistent. Most correlations between performance in selection and performance in assessments during training were positive, with the exception that performance in the CV was usually inversely correlated with subsequent assessments. The referee reports and total selection scores were predictive of performance in the major end of term work-based assessments.

This study reviewed many factors that affect selection processes and outcomes. Several of these—such as the role of procedural justice, identification of desired attributes of trainees, and the validity, reliability, fairness, and acceptability of selection instruments and protocols—implicitly framed RACS GS selection. However, to maximise the effectiveness of selection instruments and protocols, these and other relevant influences could be explicitly defined for the local context. The findings regarding the predictive capacity of the selection instruments—

particularly for the referee reports and the interviews—differ from some other studies, but highlight that the implementation and content of selection instruments are key aspects affecting their performance. Combining scores from multiple instruments reduces the influence of any individual selection instrument.

Overall, with the possible exception of the CV, the study has shown that the RACS GS selection tools are performing moderately well. However, the RACS GS selection instruments and processes could be modified to maximise their effectiveness and new, emerging selection activities could be considered.

Student Declaration

"I, Zaita Oldfield, declare that the EdD thesis entitled Making the cut: An evaluation of selection into the Royal Australasian College of Surgeons' surgical training program is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work".

Signature [redacted]

Date 27 October 2017

Acknowledgements

I wish to sincerely thank my supervisor, Associate Professor Anthony Watt, for his guidance and support. Tony has provided direction and challenged me to explore and understand additional areas of research and scholarship. Tony's positive attitude and good humour have enriched the process.

I am also extremely grateful to Professor Julian Smith who was invariably positive and patient in his role as co-supervisor. His critique of the research process and the thesis manuscript was always pertinent and encouraging. I have greatly appreciated Julian's confidence in my ability to accomplish this project.

Many thanks to the Royal Australasian College of Surgeons for supporting this project, and to the numerous colleagues who have taken an interest along the way. Countless discussions have provided me with new perspectives on my research and writing and have buoyed me at critical stages throughout the process. I greatly appreciate the practical advice and assistance that I have received. I also thank the Board in General Surgery and General Surgery administrative staff, particularly Monica Carrarini, for facilitating access to their records.

I thank my family and friends for their enduring support, especially my mother St Clair Oldfield, who has maintained a keen interest in my progress. Most importantly, I wish to express my deepest gratitude to my partner in life, Ross Mathieson and our son, Oscar Mathieson, who have been unfailingly supportive and encouraging throughout the vicissitudes of the process, and have understood and accepted the many times when I have been absent and working. I hope we will all share in the benefits of my achievements.

Table of Contents

List of Tables

List of Figures

List of abbreviations, acronyms and symbols

Abbreviation/ acronym/symbol	**Meaning**
Narrative	
ANZ	Australia and New Zealand
ANZAPS	Australian and New Zealand Association of Paediatric Surgeons
Au	Australia
BiGS	Board in General Surgery
BSET	Board of Surgical Education and Training
BSS	Basic surgical sciences examination
BST	Basic surgical training
CE	Clinical examination
CV	Curriculum vitae (and/or structured curriculum vitae)
DOPS	Direct observation of procedural skills (an assessment of performance of a clinical procedure or part of a procedure.)
ETA	End of term assessment (an assessment of performance during a clinical rotation)
GenSSE	Generic surgical science examination
GSA	General Surgeons Australia
GS	General surgery
HAR	Hospital assessment report
ID	Identification
Int	Structured multi-station interview
MCC	Minimally competent candidate
MCQ	Multiple-choice-format question
MiniCEX	Mini clinical evaluation exercise (an assessment of performance of a clinical examination)

Abbreviation/ acronym/symbol	**Meaning**
MMI	Multi-mini-interview (or multi-station mini-interview)
NZ	New Zealand
NZAGS	New Zealand Association of General Surgeons
OIB	Ordered item booklet
OSCE	Objective structured clinical examination
Paed	Paediatric surgery
pdf	Portable document format
PS	Paediatric surgery
RACS	Royal Australasian College of Surgeons
RR	Structured referee report
SET	Surgical education and training
SET1	The first year of SET
SET2	The second year of SET
SJT	Situational judgement test
SpecSSE	Specialty-specific surgical science examination
SSE	Surgical science examination
SST	Specialty surgical training
Total sel	Total selection (score)

Correlation analysis	
**	Correlation is significant at .01 (1-tailed)
*	Correlation is significant at .05 (1-tailed)
_	No results because of too little data
N and *n*	Number of trainees or number of items
r	Pearson Correlation

Abbreviation/ acronym/symbol	**Meaning**
Regression analysis	
ANOVA	Analysis of Variance / regression value
Bold type	$p < .05$
Excl	Excluded
F	Fisher's F-ratio
p	Significance (chance value)
pr	Partial correlation
r	Multiple correlation
R^2	Squared multiple correlation
spr^2	Semi-partial variance (partial correlation squared)
ß	Beta [Standardised coefficients]
t	t-test value

CHAPTER 1: INTRODUCTION

This section presents the aims of the study and provides a brief introduction to prevailing concerns in medical selection and training to place the study within a global context. It also presents a general background to surgical education and training in Australia (Au) and New Zealand (NZ), and provides a frame of reference for the Royal Australasian College of Surgeons (RACS) General Surgery (GS) training and issues underpinning the study.

This project was undertaken to satisfy the requirements for the degree of Doctor of Education, as regulated by Victoria University. The primary supervisor, Associate Professor Anthony Watt, was a full-time employee of Victoria University. The co-supervisor, Professor Julian Smith, was a Fellow of the Royal Australasian College of Surgeons (Cardiothoracic Surgery) with no responsibility for selection of GS trainees.

Study aims

The primary aim of the current study is to evaluate selection activities implemented by RACS. The study interrogates selection into the RACS' GS training program by identifying relationships between selection activities and assessments administered during the first two years of surgical training. Specifically, this study seeks to highlight whether selection scores predict scores achieved in assessments during training. Findings from this study may inform practices to support the optimisation of selection methods in recruiting and admitting applicants who have the greatest likelihood of succeeding in the RACS' surgical education and training (SET) program.

Assessment scores derive from many factors, including candidates' personal attributes and training experiences. It would be overly simplistic to assert that assessment performance is exclusively engendered by selection performance. However, exploration of the relationships between selection scores and assessment scores achieved during training will facilitate evaluation of the effectiveness of particular selection instruments and will establish how well they allow us to determine the candidates who are most likely to succeed in the RACS training program. Predictive relationships between scores in selection and subsequent assessments would support RACS' confidence that selection instruments and protocols are consistent with assessments administered during training. However, scores at the selection stage that do not correlate with subsequent assessment scores, may provide evidence to inform modifications to RACS' selection, training and assessment processes.

Global context for the study

Medical colleges and universities worldwide grapple with similar concerns regarding selection, training and assessment. Studies of admission processes have identified multiple pressures on specialist training programs (e.g., Carmichael, Westmoreland, Thomas & Patterson, 2005, Eva, Reiter, Rosenfeld & Norman, 2004, Parry, Mathers, Stevens, Parsons, Lilford, Spurgeon & Thomas, 2006, Prideaux, Roberts, Eva, Centeno, McCrorie & McManus, 2011, Roberts & Togno, 2011). Tensions arising from regulatory and financial accountabilities, combined with evolving medical, educational and social expectations, conditions, and knowledge, have affected medical training.

Taylor (2005), in discussing selection to surgical training in the United Kingdom (UK), highlights two key issues: firstly, that early recognition of appropriate candidates is becoming increasingly important, and, secondly, that "agreement on entry criteria into surgical training programmes is causing controversy" (p.1). Additional pressures on selection arise from regulatory and other external requirements, increased accountability and obligations to provide "fair and transparent" admissions systems (Emery, Bell, & Vidal Rodeiro, 2011, p. 62). Carroll, Kennedy, Traynor and Gallagher, (2009) consider the financial pressures and economic implications of 'investments' made in training, nominating advantages of better selection:

> It is increasingly important that the investment made in surgical trainees yields positive outcomes, i.e. competent consultant surgeons. To achieve this goal...surgery in particular must develop a selection and assessment system that can better discriminate between candidates on factors that are known or

> suspected to be good predictors of success in training and clinical practice (p.1544).

Many of the developments and concerns identified globally are pertinent locally, although few studies to date have focussed on selection to surgical training in the Australian or New Zealand context. The desire to identify 'known' rather than 'suspected' predictors of success in training has particularly inspired the current study.

Australian and New Zealand frame of reference for the study

In Au and NZ, similar constraints of regulation and accountability, pressures to select the 'best' candidates, and to maximise the efficiency of training and assessment within a dynamic surgical training environment impact on surgical selection and training. In response to such pressures, at the time the current study was undertaken, RACS annually reconsidered and modified selection items and protocols. Adjustments included modifying the scored elements and the scoring structures within the selection items and reviewing the proportions that each component comprised of the total selection scores. Proposed changes were considered by RACS' education committees prior to being authorised and implemented. Review of minutes from board meetings at which selection items were approved for use indicated that modifications to GS selection items occurred in all years under review in the current study ("Minutes of a meeting of the Board of Surgical Education and Training held at RACS on Friday 8th October 2010", n.d.). I was unable to find evidence of the impacts of changes to selection practices being formally reported, implying that selection outcomes were not rigorously reviewed or evaluated.

It may be assumed that adjustments were intended to improve the outcomes of GS selection—to facilitate entry of individuals most likely to succeed and to reduce the likelihood of selecting candidates who were ill-suited to surgical training and unequal to the challenges therein—and to improve the efficiency with which the selection processes were undertaken. Review of board minutes does not reveal a rationale for proposed changes, but, presumably, performance of trainees was observed and found wanting, initiating adjustments across the continuum of selection, training and assessment to address perceived shortfalls.

When it was introduced, the RACS surgical education and training (SET) program represented a major shift from the previous surgical training model. The first selection into SET occurred in 2007, with selection in subsequent years engendering refinements to the processes and content. The climate of reform accompanying the establishment of SET stimulated further changes, making review of, and modifications to, selection and assessments commonplace. Many adjustments were made to selection items apparently on the basis of conjecture. The validity and reliability of selection instruments were not tested; instead, performance of trainees was monitored via exams, work-based assessments (WBAs) and anecdotal reports. The Board in General Surgery (BiGS), responsible for implementing protocols for selection into the GS training program, may have speculated on the outcomes of changes to selection items; however, as no studies were conducted into the GS selection items, direct outcomes of the changes were unclear and unconfirmed. This is in contrast to regular reviews that were conducted of examinations undertaken during surgical training, but similar to the introduction and review of RACS' WBAs. Annual changes to selection interview questions may also have been implemented to reduce the likelihood of candidates being conferred advantages unconnected to the attributes being tested, through familiarity with the questions—by being advised of them by previous applicants or having previously (unsuccessfully) applied themselves.

When the current study commenced, RACS had selected four cohorts into SET, and it was timely to evaluate the implementation of SET selection instruments and processes. This study is the first empirical measure of the impacts of the instruments and processes used to select candidates for entry to RACS' GS training. The desire to identify performance parameters and predictors of success in training have inspired the study and it is hoped that RACS' reviews of and modifications to selection and assessments will benefit from the findings of the study.

This study appraises relationships between selection instruments and the assessments during the first two years of GS training in Au and NZ. The study investigates selection to GS training for the years 2008, 2009 and 2010. The findings will contribute to a clearer understanding of the selection process by providing insights into the relationships between selection and early assessments in GS training in Au and NZ. The study provides data to identify selection instruments that demonstrate positive links to training outcomes, and similarly identifies selection instruments that are poor predictors of future performance. GS or RACS therefore may consider modifying selection practices in the light of findings from this study.

GS Au and GS NZ shared selection practices and used the same selection instruments. Each country administered its own selection locally and had some autonomy regarding the content of selection items. Examinations and WBAs were identical for both countries. The annual number of trainees accepted into GS SET varied throughout the period of the study, although GS was always the RACS' surgical specialty with the largest number of trainees. Normally, more people apply to train as surgeons than can be accommodated in SET and there is strong competition for places. It is incumbent on RACS and surgical specialty groups such as GS, to select those who are most likely to succeed in the program.

RACS and GS gather and store data pertaining to selection and assessment—applicants' scores in selection and trainees' examination scores are held in a secure database. RACS has established a suite of examinations as effective measures of defined aspects of knowledge, skills and attributes valued in surgical training. No study, however, has yet been undertaken to ascertain whether it is possible at selection to identify the applicants who are most likely to do well in the SET program.

Data that were available to the study reflected candidates' performance in selection instruments, and trainees' performance in WBAs and examinations. This allowed comparisons to be made between trainees' performance in selection and their performance in assessments during training. Strong performance in selection, followed by strong performance in assessments implies that selection instruments are identifying those candidates who are well matched to the training program; however, high scores in selection followed by low scores in assessments would imply a mismatch between selection and assessments, likely warranting reforms to the selection and/or to the assessment processes.

In this thesis, I will contextualise this study in the Australian and New Zealand medical training environment, and will elaborate on conceptual frameworks influencing selection to surgical training, including historical influences, person–environment fit, procedural justice, and Bloom's taxonomy. I will present key issues in surgical training and selection, attributes of particular selection and assessment instruments and processes. I will describe the methodology for the analysis, and will present the results. In the Discussion, I will review the context for Australian and New Zealand surgical selection, training and assessment, will critically examine patterns of performance and relationships between selection scores and assessment scores, and will expand on the relevance of major themes to RACS GS selection and the roles of people in surgical selection, training and assessment. Finally, I will describe implications for practice and present recommendations.

International and local surgical training

To contextualise Australian and New Zealand surgical training, and because much of the research into selection stems from international studies, a brief synopsis of specialty training and selection models used in English–speaking countries is presented. This is followed by a summary of the Australian and New Zealand regulatory framework, the framework for basic medical and specialty training and the RACS model of governance for surgical training.

International medical specialty training and selection

Much medical training has evolved from British, European or American models, and surgical training programs share many core goals and practices; however, variations in approaches to selection, training and assessment within and between countries are now widespread. Similarities include surgical training programs being subsequent to basic medical qualifications, and course governance bodies that advocate fair selection practices—which aim to admit those candidates who are most likely to succeed in training and beyond. United Kingdom (UK), Irish, Canadian and American surgical training programs are multi–staged with two or more discrete entry processes, commencing with core, or basic surgical training, followed by specialty training. Governance, and selection instruments and protocols, however, differ between these programs. Co-ordination of selection and training varies from local to national: the UK co-ordinates training through 'deaneries' or regions (Tooke, 2008; *The trainee doctor*, 2011), Canada uses a national Canadian Resident Matching Service (CaRMS) ("Canadian Resident Matching Service - Service canadien de Jumelage des résidents | CaRMS", 2015), while the United States of America uses a central Electronic Residency Matching Services (ERAS) ("Physicians and Surgeons: Occupational Outlook Handbook: U.S. Bureau of Labor Statistics", 2015; Association of American Medical Colleges, 2002).

Selection to surgical training is an area of particular disparity. Although most programs advocate the importance of objective, accountable selection practices, and many studies have been undertaken to review the effectiveness of selection instruments, there is little consensus regarding 'best practice'. A plethora of selection instruments and protocols is implemented, including: statements of self–promotion; recommendations from others—such as dean's letters, prior medical school performance; or academic achievement—such as grade point average (GPA) and 'honor' societies; formal tests—such as the United States Medical Licensing Examination (USMLE), the United Kingdom Clinical Aptitude Test (UKCAT) and situational judgement tests (SJTs); and interviews—including unstructured, semi-structured or multi–station interviews (Roberts, Khanna, Rigby, Bartle, Llewellyn & Gustavs et al., 2017). Applications in the UK include personal statements from applicants, a clinical aptitude test (UKCAT) and multi–station interviews or SJTs ("Application Form — Surgical Careers | The Royal College of Surgeons of England", 2015). Progression to specialist training in Ireland hinges on performance in core surgical training and a multi–station interview ("Applying to ST1 - ST2 - Royal College Surgeons in Ireland", 2015), while in Canada medical school transcripts, dean's letter, letters of reference, applicants' personal statements, examinations, CVs and additional documents that applicants feel will enhance their application may be considered in shortlisting candidates for interview ("Canadian Resident Matching Service - Service canadien de Jumelage des résidents | CaRMS", 2015; Pollett & Waxman, 2012). Selection to American surgical training varies between institutions, but typically include ERAS application, USMLE scores, a personal statement, a medical school transcript, letters of recommendation, a dean's letter and an interview (Association of American Medical Colleges, 2002; "Physicians and Surgeons: Occupational Outlook Handbook: U.S. Bureau of Labor Statistics", 2015; Schaverien, 2016). In addition to acknowledged selection instruments and explicit protocols, latent and implicit influences may also affect selection outcomes.

Australian and New Zealand regulatory framework for surgical training

The Australian Medical Council (AMC) and the Medical Council of New Zealand (MCNZ) work collaboratively to regulate all phases of medical and specialty training in Au and NZ. The AMC and MCNZ formally accredit medical education providers and their programs through assessment against predetermined standards (*AMC - Accreditation and recognition, 2017*). RACS has been accredited to deliver surgical training by the MCNZ since 1997, and by the AMC since 2001 (Batten, 2017).

Primary medical training in Australia and New Zealand

Training in Au and NZ, leading to a qualification that permits the holder to seek general registration as a medical practitioner—that is, to practise medicine—is undertaken at university. This stage of training is known as 'primary' medical education ("Australian Medical Council » Assessing primary medical education", 2017). Twenty Australian and two New Zealand universities ("Australian Medical Council » Accredited medical schools", 2017) are accredited to provide primary medical education. Some universities offer programs to which students may apply directly from secondary school—these are known as 'direct entry' or 'undergraduate' programs. Other programs require entrants to have earned a degree in another discipline prior to applying—these are known as 'graduate entry' or 'professional entry' programs' (*Medical*

Education in Australia and New Zealand, n.d.; “Becoming a Doctor”, 2017). Primary medical training combines theory and practical components, with the most significant clinical exposure occurring during the later years. It typically takes students four to six years to complete a university medical course. In Au and NZ, selection to primary medical training follows similar processes to international models: direct entry programs usually consider secondary school results and the Undergraduate Medicine and Health Sciences Admission Test (UMAT); graduate entry programs usually consider one or more of: performance in the Graduate Australian Medical School Admissions Test (GAMSAT), GPA in a qualifying degree, performance in a portfolio, special application, supplementary form, autobiographical statement and performance in an interview (Graduate Entry Medical Schools Admission System, 2015).

On completing a medical degree, graduates are provisionally registered to practise medicine and enter the workforce as ‘interns’—also known as postgraduate year 1 (PGY1) doctors. This 12-month component of training is usually undertaken in public hospitals. Interns undertake a series of work ‘rotations’ in a range of clinical environments, gaining experience in different medical specialties and a grounding for subsequent specialist training. Internship includes rotations in emergency medical care, general medicine, and surgery. Upon successfully completion of internship, doctors receive general medical registration through the Medical Board of Australia (MBA) (“Medical Board of Australia – Registration”, 2017; “Australian Health Practitioner Regulation Agency - Registration Process”, 2017).

The Royal Australasian College of Surgeons (RACS) surgical education and training (SET) program

The Royal Australasian College of Surgeons (RACS) is responsible for all accredited surgical training throughout Au and NZ in nine surgical specialties: Cardiothoracic Surgery; General Surgery; Neurosurgery; Orthopaedic Surgery; Otolaryngology Head and Neck Surgery; Paediatric Surgery; Plastic and Reconstructive Surgery; Urology; and Vascular Surgery. These nine surgical specialty groups train surgeons within the RACS’ surgical education and training

(SET) programs. Specialty training boards, convened for each of the nine specialties, manage specialty training semi-autonomously, under the governance of RACS. Some specialties administer Australian and New Zealand training separately, others deliver bi-national training. Australia's population is roughly six times that of New Zealand ("2071.0 - Census of Population and Housing: Reflecting Australia - Stories from the Census, 2016", 2017; "Population clock", 2015), with this difference reflected in the proportional representation of Australian and New Zealand surgical trainees.

Such close partnerships between countries as those between Au and NZ are uncommon in global medical specialty training. However, similarities in Australian and New Zealand's approaches to specialty training are further corroborated when it is noted that several other medical specialty colleges likewise co-ordinate joint Australian and NZ training (for example: the Australasian College for Emergency Medicine (ACEM), the Australian and NZ College of Anaesthetists (ANZCA), the Royal Australasian College of Dental Surgeons (RACDS), the Royal Australian and New Zealand College of Obstetricians and Gynaecologists (RANZCOG), and the Royal Australian and New Zealand College of Ophthalmologists (RANZCO), (*Specialist Medical Colleges*, 2017).

RACS' position as sole authority accredited to deliver surgical training for two countries is somewhat unusual, as is the RACS' SET model of streamed specialty training (Collins, Civil, Sugrue, Balogh & Chehade, 2008). The introduction of the RACS' SET program in 2007 marked a departure from a two–stage surgical training system to a 'seamless' program. The two–stage program had required aspiring surgeons to apply for training at least twice—firstly, for selection into generic, non-specialised 'basic' surgical training (BST), and secondly, for 'advanced' training in one of the nine surgical specialties. SET introduced a single entry point via which trainees were selected to train in their specialty of choice at the outset. SET also emphasised concepts of competency-based training, integrating formative, WBAs with summative assessments, and the devolution of numerous training responsibilities to specialty training boards which were accountable to RACS (AMC website; Collins, Gough, Civil & Stitz, 2007). Most SET training continued to occur in clinical settings (usually public hospitals), with trainees working and training in series of six–month placements—or training ‘posts’—under the tutelage of consultant surgeons, throughout the five- or six-year specialty programs. In contrast to other models, comprising arbitrary numbers of training years concluding with examinations, SET trainees’ progress was marked by formative assessments of observed clinical performance in identified competencies, as well as summative assessments of knowledge, throughout training (*Royal Australasian College of Surgeons, 2007 Executive Summary*, 2007; Collins et al., 2007). Initially, progress through SET was concomitant with time spent in the program, however, as SET matured, competency-based training was recognised for its potential to unlock the nexus between progression and time.

In Au and NZ, doctors could apply to RACS’ surgical training programs two or more years after their graduation from university with a primary medical degree—that is, from postgraduate year 2 (PGY2); surgical training then typically took five to six years to complete. Selection to surgical training is undertaken annually and is highly competitive, with the applicant cohort comprising an elite group of closely matched, highly motivated and skilled individuals. At the time of this study, the nine RACS’ surgical specialties all used variations of the same three selection tools—a Curriculum Vitae (CV), structured referee reports (RR) and interview (Int) —to discriminate between applicants, although the effectiveness of these selection tools had not been confirmed in the local context.

With the introduction of SET, RACS sought to ensure that the nine surgical specialties under its auspices used selection methods that were fair and consistent. All applicants to a surgical specialty undertook the same selection processes, with no scope for external influences or personal recommendations (other than via structured referee reports) to sway selection panels (Collins et al., 2008). There does not seem to have been any hesitation by RACS to initiate this new model for surgical education and training. Perhaps Australians and New Zealanders, having relatively recently instituted formal surgical training, and culturally being amenable to social mobility and ‘a fair go’, have been particularly receptive to egalitarian methods of selection, training and assessment and to judging performance over reputation and rewarding diligence and ability over pretension (Argy, 2006; Clark & Maas, 2013).

Interlinked with an increasing prominence of external accountability to bodies such as the AMC and MCNZ, was a seminal report into trainee selection, conducted by the Commonwealth Department of Health and Family Services (1998). This report—commonly known as *The Brennan Report*—emphasised the importance of using transparent, objective and quantifiable selection criteria. RACS endorsed the implementation of such selection and assessment methods, which were regarded as providing better sources of information for decisions and were considered to be 'defensible' should decisions or outcomes be challenged (Commonwealth Department of Health and Family Services, 1998). In conjunction with accountability and objectivity, RACS prioritised congruity in selection, training and assessment across its nine surgical specialties. Since 2007, all RACS' surgical specialties used the same three selection instruments and all trainees, regardless of specialty, undertook a core set of assessments before being certified as Fellows of RACS. However, flexibility and specialisation in the content and implementation of the selection instruments and in additional assessments during training, have tempered requirements for uniformity (Collins et al., 2007; Collins et al., 2008). For example, the RACS Board of Surgical Education and Training (BSET), responsible for regulating implementation of SET, specified proportional weighting ranges for the selection activities. Each surgical specialty accordingly allocated a weighting to each of the selection items within the agreed ranges. All RACS' specialties published their selection and training requirements online for the period of the current study and continue to do so ("General Surgeons Australia – Selection", 2016; "Become a Trainee" New Zealand Association of General Surgeons", 2016). RACS was commended for putting in "enormous efforts to update and upgrade [its] practices" (Commonwealth Department of Health and Family Services, 1998, p. 68). The RACS impartial, objective approach contrasts with arbitrary, ambiguous and inconsistent selection and training practices that once favoured a privileged few at the expense of the majority.

Governance of RACS surgical training

RACS governance structures are hierarchical. RACS governs training programs in nine surgical specialties through the RACS' Board of SET (BSET). The BSET, comprised primarily of specialty training board chairs, stipulated selection, training and assessment parameters. The surgical specialties had flexibility within these parameters to implement specialty-specific content and procedures. Proposed changes to selection processes were presented annually to the BSET for discussion and authorisation. The BSET, in turn, was accountable to RACS' Education Board and, ultimately, to RACS' Council; this multi–staged review and endorsement of SET selection and assessment practices fostered accountability in governance of surgical training.

Prior to the introduction of SET, in 2007/08 (Collins et al., 2008), GS conducted selection and training with considerable autonomy. However, with SET, the surgical specialty groups became more accountable to RACS regarding their surgical training programs. In SET, trainees were selected directly into one of the nine RACS surgical specialties (Martin, Blennerhassett, Hardman & Mundy, 2009), and advancement through SET was intended to reflect trainee competence rather than time spent in training (Birks & Palermo, 2008). Selection in early 2007, of trainees to commence SET in January 2008, was the first phase of the new training program. Although obliged to use the CV, RR and Int in selection each year, GS could designate the proportion of the score allocated to each selection instrument, the individual scores achievable with each item, and the way the selection instruments were implemented.

Selection, training and assessment in SET

Selection to RACS' surgical training programs occurs annually. The nine surgical specialties implement designated selection practices to determine who will be admitted to surgical training. Selection is staged through the first six months of each year, with training commencing in January of the subsequent year. The number of trainees selected varies each year, with the intake dependent on the number of training positions available. The number of available training positions reflects the number of extant trainees who vacate training positions—by progression through the program or by leaving the program. GS is the RACS specialty with the largest annual intake of trainees, however, at the time of this study, attrition from GS, brought about by trainees transferring to other surgical specialties, was a concern. Reasons for trainee dissatisfaction with GS training were not investigated, however, anecdotally, it was deduced that some trainees considered GS to be a stepping stone to their preferred specialty.

During SET, trainees concurrently work and train in hospitals. They are allocated to a series of training 'posts', supervised by consultant surgeons. Trainees gain clinical skills through ad hoc and planned practical experiences as well as by participating in structured training sessions. Assessment is via WBAs and summative examinations throughout the training program. The final assessment, the Fellowship examination, must be achieved before trainees may graduate as Fellows of RACS.

RACS SET training draws on a framework of nine competencies, defined by RACS in 2003, describing procedural and non-technical skills deemed essential for surgeons and surgical trainees. The nine RACS competencies comprise collaboration and teamwork; communication; health advocacy; judgement and clinical decision making; management and leadership; medical expertise; professionalism; scholarship and teaching; and technical expertise (Watters & Civil, 2011; "Competencies | Royal Australasian College of Surgeons", 2017).

RACS engages in establishing and instituting optimal practices to meet the challenges of selecting, training and assessing surgical trainees. In March 2008, for example, RACS hosted an international conference on surgical education and training, with a major goal of reaching an agreement regarding principles for selecting surgical trainees. Incorporating this and other local and international dialogue, RACS and the surgical specialties regularly reviewed and modified selection and examination instruments, policies and processes. In introducing more accountability to the selection processes, RACS reviewed models of selection in use worldwide, but decided that the unique circumstances of surgical training in Au and NZ warranted tailored solutions. The current study provides local data to inform decisions about Australian and New Zealand best–practice selection into surgical training.

This Introduction has presented the aims of the study and identified key global concerns in medical selection and training. It has also presented a general background to surgical education and training in Au and NZ, and provided a frame of reference for RACS' GS training and issues underpinning the study. These themes are developed in the Literature Review.

CHAPTER 2: LITERATURE REVIEW

Part 1 of this Literature Review presents an historical frame of reference for surgical training and three conceptual frameworks to contextualise the study and findings. Part 2 of the Literature Review, explores key issues in surgical training and selection: the goals of surgical training and selection, competition and stakes, identification of desired attributes of surgeons and trainees, aptitude and learning, cognitive, visuospatial and psychomotor skills in selection, and diversity. Part 3 identifies attributes of selection and assessment instruments and processes. These include validity, reliability, feasibility, acceptability, bias and fairness. In Part 4, I discuss selection instruments and protocols in general, and examine in more detail some prevalent selection instruments, including those adopted by RACS. In Part 5, I present a sample of studies that have used methodologies comparable to methods employed in the current study.

Part 1 Historical context and conceptual frameworks

An historical context for medical training and selection practices identifies some historical influences on surgical training and selection.

Historical influences on surgical training and selection

Initially, European approaches to surgical practice and training were ad hoc and disorganised. Until the 1800s, changes in medical education occurred “slowly and almost imperceptibly” (Warren, 1951, p. 304). However, with the development of empirical methods, medical knowledge, and the ways in which it was imparted, advanced.

Prior to the Reformation in the 1500s, barbers performed bleeding and some surgical procedures as counterparts to clerics’ and apothecaries’ care of the sick. As barbers’ healing and surgical roles expanded they established the Barber-Surgeons’ Company in 1540, the Surgeons’ Company in 1745 and, ultimately, the Royal College of Surgeons in 1800 (Warren, 1951).

From the 1500s, to qualify to practise medicine, a proficiency in Greek and Latin, was "almost sufficient in itself" (Warren, 1951, p. 309), as classical scholars attained medical knowledge by studying Greek (Warren). By the mid-1600s medical training, through lectures, could be undertaken at some universities, although they were often lax in assessment. According to Wittie (1651) as quoted by Warren: "in many Universities, although Physick be diligently taught in their publique Lectures, yet in conferring these degrees they are too carelesse, denying them to few or none" (Warren, p. 304).

During the 1700s multiple avenues—apprenticeship, university and hospital-based training, undertaken independently or together—could lead to medical practice. Warren, (1951) tells us that many an aspiring doctor "became an apprentice to his employer and a formal legal document was drawn up" (p. 305). On completing a five- to seven-year apprenticeship, the student received a certificate stating that "he had completed his training satisfactorily" (p. 305). Access to an apprenticeship was at least partly based on a capacity to pay. Abraham (1933), in Warren, describes Dr. John Fothergill's (1712–1780) terms of indenture:

> It was decreed that he 'his master well and faithfully shall serve; his secrets shall keep; taverns he shall not haunt; at dice, cards, tables, bowls, or any other unlawful game he shall not play.' In return his master undertook to teach him 'the art, trade, mystery or occupation of an apothecary,' and provide him with 'sufficient and enough of meat, drink, washing and lodging.' Fothergill's father had to pay £50 for these considerations.
>
> In the master's absence the apprentice visited the patients and at other times helped his master by writing out the prescriptions and dispensing them. Meanwhile he was expected to read as many books as possible and generally acquire a thorough knowledge of medicine (p. 305).

According to Wall (1937), as cited in Warren, (1951), during the late 1700s apprentice surgeons also faced a plethora of examinations:

> The Surgeons' Company [precursor of the Royal College of Surgeons] help [sic] many different examinations. The surgeon's apprentice had to pass a preliminary examination in Latin…before they could be bound to seven years' servitude; on completion of which they were examined in surgery, the internal speculation of the natural causes and remedies of all manner of infirmities or diseases, etc. (p. 309).

At this time, British and European universities conferred medical degrees ostensibly via thesis, although in practice, verification continued to be slipshod. Warren (1951) tells us that while it was possible for a degree to be "bought for a few shillings" (p. 306), it was also the case that Fellows of the Royal College of Physicians were able "by influence to prevent persons of insufficient attainments from admission to degrees" (p. 306). Lectures and texts—the main forms of medical training—were mostly in Latin.

Although clinical teaching was part of neither Oxford nor Cambridge University medical training, surgery was taught in some hospitals through observation, practice and lectures, all of which were paid for by the apprentice students. Students completing surgical training were issued "a certificate signed by each surgeon … stating that they had worked diligently" (Warren, 1951, p. 308). Joseph Warner's 1792 account of surgical teaching at Guy's Hospital in London, as cited in Warren (1951), highlights the following:

> 'Each surgeon was permitted to receive four pupils and four dressers at a time, inclusive of apprentices.' The apprentices were the most superior, then the dressers and finally the pupils. The pupils originally had to bring certificates of their apprenticeship, 'but now they only bring their money. … The pupils' business is only to look on, and to make such an enquiry as he shall choose of the surgeon who is then attending. … it is the business of the surgery man to acquaint the pupils with the intention [to open a body for examination]. …It is the business of the surgery man to make them acquainted with every accident immediately on its entry.' … For these privileges the pupil paid 24 guineas per

> annum or 18 for six months. This money was shared amongst the surgeons and apothecaries to the hospital. The dressers, who paid £50 per annum direct to one surgeon, became the responsibility of that surgeon and took a more active part than 'looking on,' as also did the apprentices who paid from £250 to £1,000 to their master….
>
> At the same time as this clinical teaching was carried on lectures were given. … 'The fee for [anatomy] lectures and for the dissecting room is twelve guineas. There are lectures read every morning at half-past seven on Midwifery. …At ten o'clock Mr. Babington the apothecary, gives a lecture in Chymistry. Those mornings that pass without the lecture in Chymistry, Dr. Saunders supplys with one on the Practice of Physic. The anatomical lectures are every day from one o'clock until three o'clock.' The fee for each course of lectures except the anatomical was 10 guineas (pp. 307–308).

During the 19th century medical education became more stable and uniform, and "the classical works assumed their proper importance as companions of medicine, rather than the masters" (Warren, 1951, p. 310). Hospital outpatient teaching was introduced early in the 19th century.

In northern America at this time there was no formal training—all surgeons were either self–trained or they apprenticed themselves to a 'master' (Cameron, 1997). This changed in the late 19th century when Sir William Halsted championed a German model of university-sponsored, hospital-based residency training which emphasised learning with graded responsibility (Cameron, 1997). Training comprised "an initial a stage of observation, followed by increased participation in surgical procedures under close supervision" (Wanzel, Ward & Reznick, 2002, p. 604) and focussed attention on the student (or resident) rather than on a particular professor. (Wanzel et al., 2002) report on the pervasiveness of this system, that "was adopted widely at that time and remains the cornerstone of surgical training programs today" (p. 597).

Critics of the Halstedian model of residency training have commented on its ‘pyramidal’ nature, for although several residents commenced training each year, “half were only permitted to train for 1 year, and few completed a full course of training” (Bell, Banker, Rhodes, Biester & Lewis, 2007, p. 811). In addition, the pre-eminence of the Halstedian model has been challenged by Rutkow (2013) who proposes that simultaneously, in New York, “a curious breed of medical schools [arose with] surgical-oriented curricula” (p. 1130) and that these postgraduate medical schools strongly influenced surgical training by offering short, practical courses for medical graduates who wanted to specialise in particular branches of medicine. Both the Halstedian and postgraduate models influenced surgical training in America in the early 20th century, concurrent with similar changes elsewhere in the world.

Geffen (2014) suggests that by the 1950s, medical education in the UK had evolved from “a chaotic mix of institutions and practices” (p. S19) to a two–phased system combining academic and clinical components. Students were initially trained in university-based medical schools and subsequently in academic departments in teaching hospitals. Australian medical training also followed this model with medical students being selected on the basis of their secondary school performance (Geffen, 2014). It was generally assumed that high marks were an indication of likely success in the medical professions.

Although at Australia’s Federation—in 1901—there was little or no Australian postgraduate medical study or specialisation, by the 1920s, as Storey recounts, “there was general agreement that both training and accreditation were haphazard and required both clarity and uniformity” (Storey, 2014, p. S26), with calls for “extended study, additional hospital experience and the selective influence of a stiff test” (Storey, 2014, p. S26)

RACS was established in 1927 (Beasley, 2002; Geffen, 2014) “to promote the art and science of surgery” (Syme, 1928, p. 488) admitted candidates to Fellowship, by examination. By the mid 1940s RACS “required a candidate to possess a Primary examination, to undergo a formal period of training and then to pass a Final examination” (Beasley, 2002, p. 85). However, as recently as 1954 there were “many well–established surgeons who … were not prepared to submit to formal examination [to gain RACS Fellowship]" (Beasley, 2002, p. 88). The examination interview was perceived by some as discriminatory; Miller (n.d.), as quoted by Beasley (2002) criticised the examination interview as:

> a rather unpleasant and unsatisfactory procedure, in which the candidate confronted the whole Court [of Examiners] across a table and could be bombarded with questions which were not always appropriate, and [which] were at times unsympathetic (p. 88).

The level of interest and consequent changes to surgical training and assessment were such that between 1926 and 1958 "there had been fourteen ways of becoming a FRACS" (Beasley, 2002, p. 89). In the late 1970s "basic surgical training programmes" were again restructured and a "register of basic trainees" was discussed, however in the early 1980s it was felt that "to place the name of a would-be basic trainee on a register … would confer on that person a status which might be misinterpreted as an obligation on the part of the College … [for] concession or recognition" (p. 139).

In 1969 "there was broad [College Council] agreement on the need for planned programmes in specialist surgical training under the supervision of an appropriate authority" (Beasley, 2002, p. 149). This manifested in the 1970s as a two–phase training program culminating in an exit Fellowship examination. This two–phase system of basic surgical training (BST) and 'advanced' specialty training (AST) necessitated two selection processes, as success in BST did not guarantee entry to AST. Surgical specialties devised their own selection methods, which reflected specialty priorities and varied in rigour:

> In New Zealand prospective orthopaedic trainees were 'vetted' at a weekend retreat.... This worked well. Trainees were judged on something more than a brief interview, and in turn came to know some of the older surgeons.... I can recall commending the idea to a general surgical colleague of mine who was involved in his own selection process. 'I see,' he commented, 'you fellows judge

> your trainees on how they hold their glass in the evenings, as well as on their reports' (Beasley, p. 152).

Geffen (2014) reports that in the late 1990s, revisions to curriculum content and expanded admission procedures, "elicited considerable controversy … particularly among surgeons" (p. S20) who considered that such revisions eroded admission standards and neglected scientific knowledge. Presumably, these surgeons felt that success in training was predicated on trainees' 'scientific knowledge' at selection. However, these changes were at least in part a response to new influences such as the Canadian CanMEDS Physician Competency Framework, which described the knowledge, skills and abilities of specialists in seven domains—as medical experts, communicators, collaborators, managers, health advocates, scholars and professionals ("Royal College of Physicians and Surgeons of Canada: CanMEDS Framework", 2016). Wanzel et al. (2002) summarise some of the changes, identifying that "medicine as a profession has become more systematized in the recent past" (p. 649). Iobst, Sherbino, ten Cate, Richardson, Dath, and Swing, (2010) concur: "with the introduction of *Tomorrow's Doctors* [UK Medical Council standards for medical education] in 1993, medical education began the transition from a time- and process-based system to a competency-based training framework" (p. 651). The competency-based training model has, to varying degrees, become adopted in specialty medical education in many countries through the early part of the 21st century with ensuing changes to selection methods.

It is implicit—when the onus was on an aspiring surgeon to find a master and to pay to become their apprentice—that deep pockets and an existing relationship with the master would stand an aspirant in good stead when the master was choosing whom to accept. With the systematisation of medical training, and selection being controlled by organisations placing increasing emphasis on objective methods, selection instruments such as letters of recommendation and referee reports—proxies for personal 'introductions'—may be seen as remnants of arbitrary arrangements when 'who you knew' could sway selectors' judgement and secure your appointment. In a similar vein, applicants seeking opportunities to 'introduce' themselves by presenting themselves in their best light, showing what they can do, and interacting with selectors (Burgess, Roberts, Clark & Mossman, 2014) nowadays rely on instruments like the CV and the interview to convey this personal information where once they would have made representations to a master or their associates.

Conceptual frameworks

To contextualise selection to surgical training in Au and NZ, to provide foci for understanding this study's results and to inform discussions regarding practical applications of the findings, I have identified three conceptual frameworks. Literature regarding surgical selection tends to concentrate on functional aspects and implications for practice in specific environments, resulting in a lack of theoretical models for selection to surgical training. Conceptual frameworks describe the main phenomena studied—the key factors, constructs or variables—and the presumed relationships among them (Miles & Huberman, 1994). For the purposes of this review, three identified frameworks have been gleaned from the fields of human resources and education. The three conceptual frameworks presented are: person–environment fit theory; procedural justice theory, and Blooms taxonomy.

Person–environment fit theory

Medical training does not occur in a vacuum; it is influenced by prevailing social norms—as these evolve, so does the content and presentation of training. Consequent on changes to training, are changes to selection principles and processes—the nature of training influences selection methods. Some selection processes will be more attuned to the idiosyncrasies of ensuing training programs than others. Aligning selection with training—to rank highest those candidates most suited to a specific training program—is a challenge. Mismatches may result in the "appointment of trainees who struggle with a particular curriculum or training culture" (Bell, Fann, Morrison & Lisk, 2012, p. 23). Kelz, Mullen, Kaiser, Pray, Shai and Drebin (2010), from their study of resident attrition, suggest that "it is plausible that to reduce attrition, programs need to work to 'match' trainees that have characteristics that are compatible with their type of program" (p. 537).

In organisational behaviour research, person–environment fit describes the "congruence, match, or similarity between the person and environment" (Edwards, 2008, p. 168). This theory underscores the notion that the better the fit between the attributes of an individual and the characteristics of a vocation—between the needs and abilities of the individual and organisational demands and rewards—the greater the job satisfaction and the more likely that the individual will meet the required performance standards and be retained. When this is extrapolated to the realm of specialty training, it suggests that the greater the concordance between a trainee and a training program, the more likely that the trainee will complete training.

Go, Klaassen and Chamberlain (2012) reflect on how this may influence selection, suggesting that candidates' "non-academic qualities and 'fit' within a programme are playing increasingly significant roles in recruitment" (p. 498). This is a two–way process in which organisations identify and assess candidate attributes and candidates are able to identify features of training programs. Kelz et al. (2010) recommend that recruitment information clearly identifies both positive and negative aspects of a program's educational style and culture—addressing "the realities of the surgical training environment" (p. 537). In their study of German, UK and Swiss surgical training, von Websky, Oberkofler, Rufibach, Raptis, Lehmann, Hahnloser and Clavien (2012) similarly identify that the management of trainee expectations contributes to trainee satisfaction and retention—they recommend clarifying expectations through documented and structured training curricula to increase trainee satisfaction. The selection process itself contributes to expectation management and exchange of information. Burgess et al. (2014), in their study of selection into general practice training, argue that "properly conducted selection systems are in the best interest of both the candidate and the organization" (p. 3) and that components of the selection process ought to provide "insight and understanding of what is required to work in general practice" (p. 4). A novel example is cited by Seabott, Smith, Alseidi and Thirlby (2012), who describe a "candidate-centered" (p. 803) approach to interviewing, in which applicants each individually spend a day observing and interacting with trainers and residents as they undertake their normal clinical and surgical activities. This "working interview" (p. 805) explicitly encourages open disclosure of information about the program and offers both 'interviewers' and candidates opportunities to observe and interact in authentic clinical environments.

Singletary (2010) questions whether a lack of awareness of "the challenges of a surgical residency" (p. 365) contributes to trainee attrition; he puts the onus on candidates to inform themselves and accurately judge their fitness for surgical training before applying, "It is hard to believe that students interested in a surgical residency are not aware of the challenges that await them. …It is certainly possible that some students opting for a surgical residency may overestimate their ability to adapt to the demanding schedule" (p. 365). A counter view is presented by Buhr, Gröne and Ritz (2012), who highlight the importance of two–way interactions between trainees and trainers, suggesting that "good training in surgery always depends on two people: a motivated trainer and a resident who wants to be trained. Personal commitment is always required on both sides" (p. 808). The quality of people's interactions and extent of their commitment may reflect the affinity between the culture of the organisation and the individuals functioning within it.

The expectations that employers and employees have of each other have been described as a 'psychological contract', or set of beliefs that are held to be important by each party. These expectations and assumptions, often unarticulated, can be crucial to the person–environment fit (Armstrong, 2006; Rousseau & Greller, 1994; Sims, 1994). Sims (1994) proposes that "a balanced psychological contract is necessary for a continuing, harmonious relationship between the employee and the organization. However, the violation of the psychological contract can signal to the participants that the parties no longer share (or never shared) a common set of values or goals" (p. 375). Armstrong (2006) recommends that it is incumbent on managers to "manage expectations [by] clarifying what they believe employees should achieve, the competencies they should possess and the values they should uphold" (p. 227). This viewpoint has great resonance with surgical training.

The importance of a harmonious person–environment fit intensifies when pressures are brought to bear on organisations and individuals. Armstrong (2006) observes that "leaner organizations may make greater demands on employees and are less likely to tolerate people who no longer precisely fit their requirements" (p. 231). Constraints such as a shorter working week are placing pressures on surgical training—in effect making training 'leaner'. If Armstrong (2006) is correct, the precision of the fit between trainees and training programs becomes increasingly important.

Awareness of the person–environment fit framework as it applies to selection to surgical training enhances our understanding of the two–way nature of selection—that applicants are 'selecting' a surgical career as much as surgical specialties are selecting trainees. This understanding places greater emphasis on candidates' knowledge of and judgements about surgical training programs than is commonly understood in Au and NZ. It also highlights the importance of establishing that the range of skills and attributes sought in trainees will suit them to careers in surgery.

Procedural justice theory in selection

Procedural justice—the fairness of the process by which outcomes are determined (Lind & Tyler, 1988) —is today implicit in most recruitment, selection and assessment protocols. Procedural justice has been said to affect candidates, organisations, and the quality of information gathered. Organisations conducting selection to surgical training are responsible for implementing procedures that ideally, conform to principles of procedural justice; these organisations conduct their selection practices with varying degrees of autonomy, accountability and proficiency.

The seminal study of trainee selection in Australian medical colleges—commonly known as *The Brennan Report* (Commonwealth Department of Health and Family Services, 1998)—

found marked differences between selection processes implemented by the major Australian specialty medical colleges; the report emphasised that some graduates thought selection processes were unfair. To address this, *The Brennan Report* (1998) set out principles and a framework for trainee selection into specialist education and training. The framework fosters transparency and procedural justice, stipulating 14 requirements to which colleges must adhere when selecting into their training programs. A summary of the framework is presented in Figure 1, (Commonwealth Department of Health and Family Services, 1998, pp. 81–97). The 'Brennan principles' are still evident throughout Australian and New Zealand specialty selection processes and continue to underpin RACS' selection to surgical training.

Statement of principles	There should be a clear statement of principles which underpin the selection process.
Eligibility criteria	There should be a clear statement of eligibility to apply for, and be selected for, training.
Advertising	There should be national awareness of opportunity for all eligible candidates.
Limits to the number of training positions	Quotas, if applicable, and limits relating to other factors, such as the number of training
References	Referees' reports should be Proforma with a view to achieving, objectivity, comparability and quantification
The selection committee	The Committee should have the confidence of the candidate, the profession and the community. It should be prepared to be held accountable for their decisions with the size of the Committee proportional to the task. They should be prepared for their processes and decisions to be reviewed in other forums. The selection process should be valid, reliable and feasible with evaluation built into the process.
Selection criteria	The selection criteria should be documented and published. To the greatest extent possible they should be objective and quantifiable.

Conduct of the interview	The interview should be objective and free of bias.
Selection	The selection process should be based on the published criteria and the principles of the College concerned whilst also being capable of standing up to external scrutiny.
Ranking	The Selection Committee should score and rank candidates using the tools presented earlier.
Documentation	Adequate documentation enables external scrutiny, audit and evaluation of the selection process. It should enable accurate reconstruction of the original detail and process.
Feedback	The principle to be followed is that candidates should be given or at least offered a frank appraisal of their standing in the eyes of those conducting the selection process.
Evaluation	The principle is that there should be a formal, regular inclusive review of the process.
Appeals	There should be a formal process for reviewing/appealing decisions in relation to selection. Applicants should have the right to appeal externally without fear of bias and be required to bear the cost of the appeal if it is unsuccessful (Colleges to bear the cost if its s successful).

Figure 1. Summary of best practice framework for trainee selection

Beasley (2002) identified trainee complaints and challenges, stemming from perceived unfair rulings at selection or dismissal from training, as major concerns for RACS. He proposed that the College must be “seen to be just in its rulings” (p. 152)—synonymous with implementing procedural justice principles. Australian and New Zealand concerns for procedural justice in selection and training are echoed elsewhere. Recognition of the importance of procedural fairness in selection to generic and specialty medical training can be seen implicitly and explicitly in regulatory requirements worldwide, although implementation of these principles varies. Regulators of specialty medical training in Britain, the General Medical Council, specified in *The Trainee Doctor* (2011) that “processes for recruitment, selection and appointment must be open, fair and effective” (p. 18) when criteria for candidate eligibility, selection processes, composition of selection panels and the appointment process were being determined.

Prior to this, also in the UK, Gough and Bell (1989) criticised the “subjective and haphazard forms of selection” (p. 975) and, more recently, Adam, Dowell and Greatrix (2011) called for medical schools to “use legitimate criteria to discriminate between applicants” (p. 1). In Canada, Robins, McInnes and Esmail (2014) refer to research conducted by Provan and Cuttress (1995) to support the view that “residency program directors favour objective data to guide selection” (p. 1 of 4). Although, when Robins et al. examined information provided by medical schools to residency program directors in support of applicants, they found that only three of seventeen Canadian medical schools included objective data, in “an interesting juxtaposition to their own admission requirements” (p. 2).

Despite general acknowledgement of the desirability of procedural justice in selection, implementation can be inconsistent. In the USA, Hern, Alter, Wills, Snoey and Simon (2013) surveyed residency applicants regarding the extent to which they were asked “potentially illegal” (p. 1546) questions during selection interviews—questions relating to, for example, marital status, family planning, age, ethnicity, religion or sexual preferences. They reported that most respondents were asked at least one potentially illegal question. Whether such questions were sanctioned prior to the interviews or were the result of impromptu probing by interviewers, this highlights the potential for procedural justice to be compromised in selection processes.

Adherence to principles of procedural justice may be a regulatory requirement, but the reasons for promoting procedural justice go beyond simply addressing trainee complaints. Peoples’ perceptions of how fairly or unfairly they were treated have been identified as affecting ongoing relationships between individuals and organisations, and procedural justice may also affect the quality of information gathered. This is particularly important when the 'currency' of selection transactions is information exchange.

Deciding who, from a pool of applicants, is most likely to succeed in an endeavour can be an uncertain process. To address this uncertainty, selectors and applicants alike rely on receiving information that is as accurate, honest and complete as possible (Breaugh, 2008; Cable & Yu, 2006). Fromm (2002) argues that uncertainty and fairness are so closely linked that it is impossible to understand one without the other. The uncertainties of selection relate to its predictive function. Klotz, Motta Veiga, Buckley and Gavin (2013) propose that because potential employees and employers cannot supply proof that they "will always fulfill the expectations of the other party when future contingencies arise, the trustworthiness that job applicants and recruiting organizations perceive in one another during pre–entry processes becomes … a proxy for such certainty" (p. S104). These perceptions of trustworthiness are meaningful as descriptors of current, observed performance and as indicators of likely future performance.

To conform to principles of procedural justice, organisations must enact, or 'model', the ethical behaviour that they expect of employees—or trainees. Recruiting organisations can provide equitable environments for interactions with applicants by using procedures that are fair—by operating consistently and without bias, by seeking and using accurate information, by allowing unfair decisions to be corrected, by encompassing all parties' needs and by acting morally and ethically (Lind & Tyler, 1988).

Some contend that the way information is gathered impacts on its ability to support judgements. Blader and Tyler (2002), for example, argue that procedural justice affects the quality of decisions and propose it as "one of the most potent influences on organizational attitudes and behaviors" (p. 108). Klotz et al. (2013) assert that "hiring organizations are interested in collecting as much information as possible about job applicants" (p. S108), and rely on the authenticity and relevance of that information when making judgements about candidates. Procedures that conform to principles of natural justice are likely to provide verifiable, authentic information. This is consistent with the position that it is beneficial for organisations to consider the benefits of fair treatment and the costs of unfair treatment when developing and implementing processes and actions (Lind & van den Bos, 2002).

As van den Bos and Miedema (2000) suggest, people need fairness when they are uncertain about things that are important to them, particularly "when they are concerned about … social interdependence and socially based identity processes" (van den Bos, Wilke and Lind, 1998, p. 1450). In high-stakes selection, candidates' future livelihoods and professional identities hinge on the fairness of judgements and the quality of the processes by which these judgements are made. Fairness is seen to "provide protection against things people are uncertain about" (van den Bos & Miedema, 2000, p. 364). For candidates and RACS, selection outcomes are uncertain. Increasing participants' perceptions of the fairness of selection processes may mitigate their concerns about the uncertainty of outcomes.

Discussing the role of 'trust' in employee recruitment and selection, Klotz et al. (2013) suggest that interactions, particularly "trust violations" (p. S115) during selection processes affect future interactions between both parties. If the influence of these initial impressions holds true for selection to surgical training, then it is crucial that these interactions are consistent with desired long–term behaviours. Relationships between individuals and surgical colleges may be life–long, continuing after trainees qualify as fellows.

Aspects of procedural justice contributing to quality decisions in selection include open access; formal, 'transparent' procedures; impartial decisions, reliable information, consistent implementation and accountability. Defining and regulating procedures increases procedural transparency and accountability, while informing interested parties also increases accountability and can boost confidence in the methods used to make decisions. Blader and Tyler (2003) link *formal* decision–making to a "fairness of procedures prescribed by … rules" and link *informal* decision making to processes originating with "particular agents of the organization" (p. 117). In practice, the degree to which selection processes and decision–making are formalised varies between organisations and with the degree of regulation or autonomy that organisations apply to 'agents', or individuals authorised to act on their behalf.

The contribution of procedural justice theory to selection in surgical training is pervasive. Many studies of selection and regulators of selection processes confirm that procedural justice plays an important role in selection outcomes and the acceptability of selection methods to candidates and selectors alike (Burgess et al., 2014). Procedural justice may be explicit or implicit, and the degree to which it is implemented varies.

Bloom's taxonomy: domains and levels

Bloom's (1956) well-known taxonomy categorises educational objectives into three domains: cognitive (knowledge-based), psychomotor (skills-based) and affective (attitudinal based). Learning in each domain is understood to progress through sequential levels of increasing complexity. Bloom's original levels in the cognitive domain—knowledge, comprehension, application, analysis, synthesis, evaluation—were revised by Anderson and Krathwohl in 2001. The revised levels—remembering, understanding, applying, analyzing, evaluating and creating—are described in more detail in Figure 2.

Bloom (1956), and Anderson and Krathwohl (2001), each describe increasingly complex cognitive activities as they progress from basic 'remembering' to 'creating' and synthesis. Others have developed activities in the psychomotor and affective domains. The intuitive logic and simplicity of Bloom's (and Anderson and Krathwohl's) model have helped this taxonomy and its modifications to dominate "most of the world's thinking of educational objectives" (Pangaro & ten Cate, 2013, p. e1200). An aspect of the taxonomy particularly relevant to surgical education is the recognition that all three domains contribute to trainee performance—that performance depends on cognitive skills, alongside other psychomotor and affective skills (Khan & Ramachandran, 2012). Understanding this suggests that skills and attitudes can, and should, be considered with knowledge (Crossley & Jolly, 2011) when contemplating the aims and objectives of training and the ensuing selection, learning and assessment activities.

Remembering	Retrieving, recognising, and recalling relevant knowledge from long-term memory.
Understanding	Constructing meaning from oral, written, and graphic messages through interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining.
Applying	Carrying out or using a procedure through executing, or implementing.
Analysing	Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure or purpose through differentiating, organizing, and attributing.
Evaluating	Making judgments based on criteria and standards through checking and critiquing.
Creating	Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure through generating, planning, or producing.

Figure 2. Revised Bloom's Taxonomy

Bloom's (1956) taxonomy supports assessment by specifying levels of knowledge, skills and attitudes that can be assigned to the acquisition of competence and stages of training. While it may be considered easier to test superficial knowledge and simple task performance, the identification of a range of increasingly complex levels in each domain provides assessors with frameworks to 'scaffold' assessment tasks and to reduce the likelihood of superficial assessments prevailing over tests of interpretation and synthesis.

Although Bloom's (1956) taxonomy is primarily considered in connection with learning and assessment during a training program, it might similarly contribute to selection—which may be considered as a form of assessment. Recognition of the three domains and their individual and collective pertinence to surgical training could assist selectors and educators to describe the content and levels of knowledge, skills and attributes required of trainees at entry and at stages throughout training.

An understanding of the relative importance of the cognitive, psychomotor, and affective domains to surgical selection and training could support development of selection, training and assessment activities, tools and measures. This composite approach contrasts with common former selection practices that were heavily weighted toward cognitive ability and/or were influenced by arbitrary criteria such as the reputation of, or a direct connection to, an applicant. Testing candidates' cognitive abilities became increasingly important as selection practices became more 'objective' and 'accountable'—perhaps partly because tests of knowledge are relatively easy to implement and measure, "measuring the measurable rather than the important" (Crossley & Jolly, 2011). In selection to medical training, measures such as GPA in previous courses and performance in examinations have represented trainees' knowledge, or cognitive achievements. The level of complexity of this knowledge is usually unspecified, but Evgeniou, Peter, Tsironi and Iyer (2013) and others suggest that assessment tends to remain at the superficial level of factual knowledge, without testing higher order understanding, interpretation, problem–solving or decision–making (Crossley & Jolly, 2011).

An advantage of using Bloom's three domains to categorise the activities and qualities sought in trainees lies in the hierarchy of levels, which could be used to define performance standards appropriate to selection and throughout surgical training. Pangaro and ten Cate (2013) agree that in frameworks, such as Bloom's, domains of competence "can be measured discretely" (p. e1202). Identifying particular qualities sought in professionals and mapping them to domains provide performance and assessment criteria, however Pangaro and ten Cate (2013) caution that although this approach "nears a fully comprehensive description of what we expect a physician to be" (p. e1203) it can also "lead to long and very detailed lists of objectives that tend to lose clarity" (p. e1203). In training and assessment, therefore, a balance is required between detailed analysis (however accurate) and a considered, but more limited, hierarchy of work-based activities, categorised into a framework. Pangaro and ten Cate (2013) emphasise that because competence in medical practice is multidimensional, Bloom's knowledge, skills and attitude domains must be "grounded in practice" (p. e1203).

The application of Bloom's (1956) taxonomy to training and assessment is extensive, but it is less commonly adopted in selection. Recognising that assessment encompassing the cognitive, psychomotor and affective domains could assist selectors to match candidates' observed performance with desired trainee qualities, at appropriate levels could enhance selection processes and outcomes.

Conceptual frameworks summary

Each of the three conceptual frameworks contributes to our understanding of selection as it is conducted today. Person–environment fit provides a sociological context, emphasising the two–way nature of selection; procedural justice theory provides a regulatory context, underscoring the accountability of selection processes that must withstand external scrutiny and Bloom's taxonomy provides an educational context, highlighting the roles of knowledge skills and attitudes. These frameworks offer meaningful insights into four facets of medical training that work together to underpin and substantiate the present study.

Part 2 Key issues in surgical training and selection

Goals of surgical training

A brief introduction to goals of specialty education and training and to the training environment will support discussion of the purpose of selection and assessment. Particularly in the last ten to twenty years, approaches to surgical training have undergone scrutiny resulting in multiple changes. Gallagher, Ritter and Satava (2003) refer to "revolutionary advances" and "questioning [of] the training paradigm that had served surgery well for a century" (p. 1525). This attention to training methods has been accompanied by reflections on the outcomes of training. Specialty surgical education and training is intended to prepare trainees to be competent, effective, ethical practitioners in unsupervised clinical practice as consultant surgeons (Bore, Munro & Powis, 2009; Carroll et al., 2009; Elfenbein, Sippel, McDonald, Watson, Scarborough & Migaly, 2015; Gallagher, O'Sullivan, Neary, Carroll, Leonard, Bunting & Traynor, 2014). White (2002), discussing orthopaedic surgical training additionally identifies leadership, research, teaching, administration and altruism as desirable outcomes for graduates. In pursuit of these training goals however, the question is equivocal whether more onus lies with trainees to learn or with the training program to teach. Bell, Fann, Morrison and Lisk, (2011) consider it is residents'—or trainees'—responsibility "to acquire the knowledge base necessary to practice [sic] competently … and to develop the qualities implicit in a professional" (p. 538), whereas others suggest that it is the responsibility of a residency program "to provide trainees with … essential skills, knowledge, and behaviours" (Elfenbein et al., 2015, p. 1098). Singh, Aggarwal, Pucher, Duisberg, Arora & Darzi, (2014) similarly propose that quality surgical training programs will "produce competent surgeons capable of delivering high standards of patient care" (p. 634). Grantcharov and Reznick (2009) not only consider it the responsibility of surgical training programs to "produce" competent professionals but to provide "a safe and pedagogically efficient environment" (p. 104) to facilitate this. While optimal training environments are likely to be specialty– and context–specific, it is noted that medical training "must adhere to the principles of medical professionalism" (Bannon, 2005, p. 70) — highlighting the fundamental imperative for consistency between training methods and optimal practitioner traits.

Goals of selection

Selection usually involves judgements about 'inclusion' or 'exclusion'; selection benefits from explicit criteria for these judgements, with selection outcomes reflecting judges' assessments of applicants' performance against these criteria. Selection is customarily used to 'filter' candidates from a preponderant group into a smaller one, limiting membership to individuals who most markedly display the desired attributes (Kulatunga-Morutzi & Norman, 2002). In the case of selection to surgical training, selection decisions reflect judgements made by authorised personnel, scoring applicants' performance in predetermined activities, as measured by selection instruments.

A key goal of selection to surgical training is to admit those who are most likely to succeed in training and beyond and to reject those who are unlikely to flourish. Selection presents challenges, several of which pertain to the uncertainties inherent in using past achievements or current performance to predict an individual's subsequent attainments. Selection is aspirational—selectors aspire to select individuals who will make the best trainees and surgeons (Carroll et al., 2009; Cleland, Dowell, McLachlan, Nicholson & Patterson, 2012; Cuschieri, Francis, Crosby & Hanna, 2001; Elfenbein et al., 2015; Gallagher, Neary, Gillen, Lane, Whelan, Tanner & Traynor, 2008; Martin, 1996; Makdisi, Takeuchi, Rodriguez, Rucinski & Wise, 2011; Thordarson, Ebramzadeh, Sangiorgio, Schnall & Patzakis, 2007); however, selection decisions are frequently expressions of speculative conjecture or inference about candidates' potential for the future. These speculations are far less conclusive than deducible outcomes would be; many programs strive to minimise the uncertainties and to employ proven methods in selection decisions (Kulatunga-Moruzi & Norman, 2001).

One proposal to formulate an effective selection process is for those responsible to "define what is required for successful trainee performance and then to systematically and objectively evaluate these attributes [in selection]" (Carroll et al., 2009, p. 1544). Although this initially appears straightforward, such actions rely on the outcomes of imprecise activities—defining relevant attributes for 'successful performance' and employing appropriate assessment instruments and techniques to assess candidates for these attributes. Additional influences, such as regulatory and social frameworks, human 'agency' and performance and the intensity of competition add complexities to selection processes. The effects of the interplay between the various components in the selection process—the goals, the instruments, their implementation, the frameworks, and people—affect selection outcomes. This also raises the notion of selection as a system (Patterson, Dowell, Nicholson, Cousans & Cleland, 2016). Figure 3 depicts some of the main influences on selection to surgical training in Au and New Zealand.

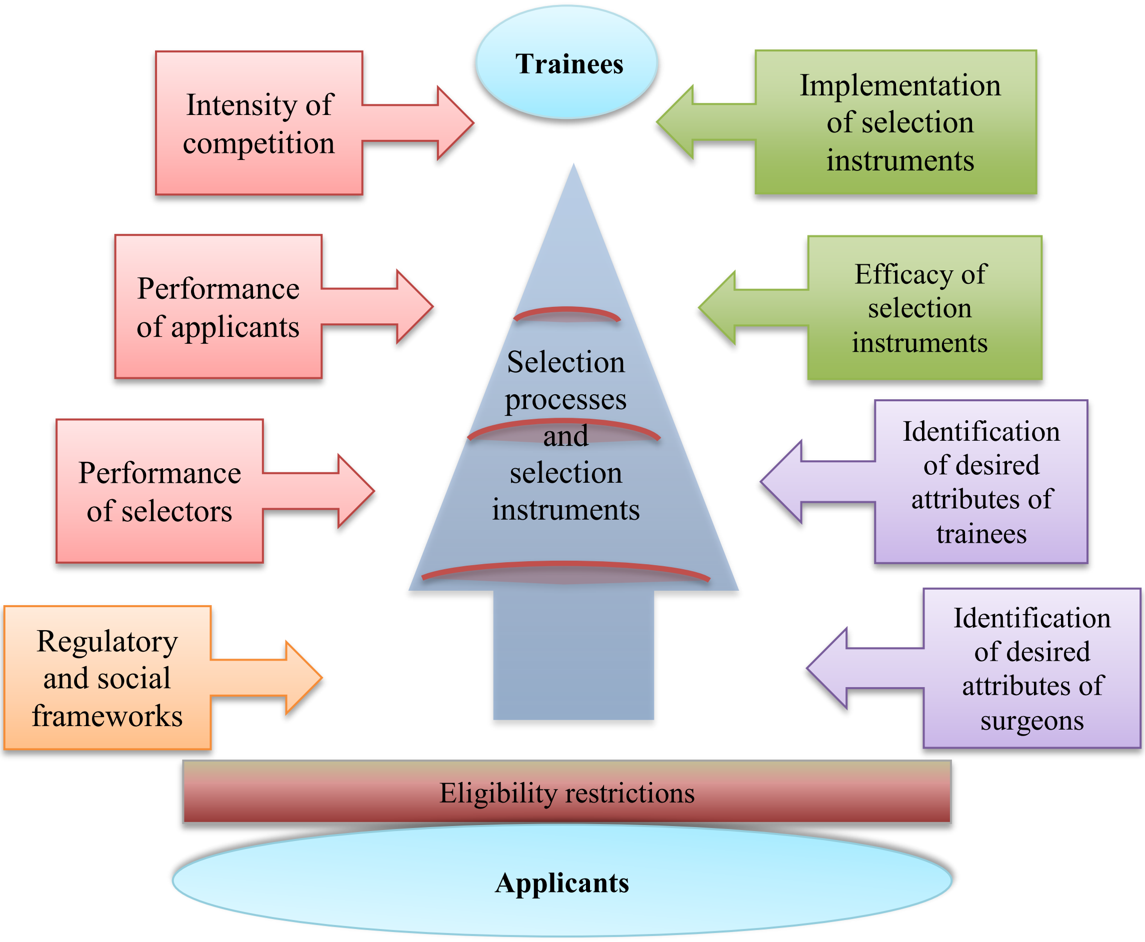

Figure 3. Influences on selection to surgical training in Australia and New Zealand

Addressing these issues in order to optimise selection outcomes and minimise the consequences of imperfect selection processes continues to tax surgical educators. Many studies explore some of these challenging facets of selection, however, few researchers consider the combined impacts of all of these issues on selection outcomes—some researchers make fewer distinctions or conflate these influences in different ways (Patterson et al., 2016). Identifying and analysing the many components of selection allows us to consider their likely influences on selection outcomes and offers the potential to maximise their positive contributions. Analysis of how they work together may result in more effective, integrated selection systems. This notion is influenced by concepts of programmatic assessment, that is:

"the quality of the programme [or system] is built on the quality of the *combinations* [my emphasis] of its building blocks" (Schuwirth & van der Vleuten, 2012).

Trainee selection has been described as "one of the most important educational responsibilities for medical school faculty" (Thordarson et al., 2007, p. 255), with the desirability of places in surgical training rendering selection intensely competitive (Carmichael et al., 2005). However, Dirschl (2002) observes the difficulty of implementing effective selection practices, and Bell et al. (2011) decry selection of residents (trainees) as "an inexact science," warning that "errors can prove costly, disruptive, and potentially damaging to training programs, and personal and professional setbacks can occur for resident applicants" (p. 534). These views indicate that selection to surgical training is considered to be vitally important and highly competitive yet its implementation is difficult and imprecise.

Focussing on the transition from medical student to surgical trainee, Makdisi et al. (2011) propose that the selection process should "evaluate candidates … to discern the qualities needed to transition successfully from medical school to residency [as] successful resident selection is critical to the process of high-quality surgical education" (p. 67). Gardner, Ritter, Paige, Ahmed, Fernandez and Dunkin, (2016) suggest that selection is context-specific, with the goal being "to identify…those who are most likely to succeed…in a specific program" (p.535). Many authors, however, take a longer-term view. Martin (1996) advocates that "each surgical training programme wishes to select the 'best and brightest' candidates: those who will ultimately develop into the best surgeons" (p. 428), while Carroll et al. (2009), considering plastic surgery training in the Republic of Ireland, more pragmatically suggest that "the aim of a selection process is to identify and select those trainees most likely to develop into competent and effective surgeons" (p. 1544). Thordarson et al. (2007) agree, suggesting that ideally, selection will identify "those who are most eligible to be trained into competent, caring, and professional doctors" (p. 255). Others have considered risks of sub-optimal selection, proposing that the goals of selection are both to identify individuals who will complete training to develop into competent specialists *and* to identify and reject those unlikely to succeed or who may become "problem" trainees or surgeons (Collins, 2007; Cone, Byrum, Payne & Smith, 2012; Gallagher, Leonard & Traynor, 2009; Roberts & Togno, 2011; Vassiliou & Feldman, 2011).

There may be a tension however, between attributes that are desirable in trainees and those that could become more important later in one's career. Cuschieri et al. (2001) touch on this when they suggest "recruitment of persons with the appropriate aptitudes to any profession underlies eventual performance" (p. 110). Thordarson et al. (2007) more directly question whether the attribute of reliability, which "probably counts more in a resident" (p. 259), would ultimately prove to be less important than "creativity" (p. 259) in a surgical career. This introduces an additional area of speculation ensuing from selection, regarding the alignment of selection criteria and performance beyond training, into independent practice.

Many authors have commented on the preferred attributes of surgeons and of trainees, formulating their perspectives by a variety of means. While there is no consensus on the composition of attributes, there are many similarities between recommendations and all agree that a mix of attributes among both practising surgeons and trainees is essential.

Competition and stakes

Competition.

Intensity of competition in selection to surgical training is affected by the number of places available, the number of candidates seeking those places, the frequency of opportunities for selection, the desirability of achieving placement and the consequences of selection outcomes. Annual selection to GS SET is inherently competitive (Farah, Winter & Smith, 2011)—all candidates are scored on their performance in selection activities, with offers for a finite number of GS SET positions being made to those who score highest. This form of competitive selection is typical of many surgical training programs (Gallagher et al., 2014; Maan, Maan, Darzi & Aggarwal, 2012; Salvatori, 2001). Competitive selection is intended to allow those best suited to enter training, however, some unintended outcomes also ensue.

Particularly at medical school, competitive selection can exclude, for example, those from disadvantaged and low socio-economic backgrounds, some ethnic and cultural groups and those with disabilities (Cleland et al., 2012; Prideaux et al., 2011). Such exclusions may reflect neither the intention of selectors, nor these aspirants' aptitude for medical training, but result from candidates' lack of access to developmental opportunities, disadvantaging them when competing against individuals from more favourable backgrounds. Competitive selection has also been criticised for encourage some candidates to lie in order to place themselves ahead of others (Young, 1997). Jefferis (2007) recommends, "the process of embarking on specialty training should be transparent and straightforward as well as competitive" (p. 1304). Gallagher et al. (2014) add that a "strong competitive selection model" (p. 303) is comprised of "objectively assessed" (p. 303) determinants.

The stakes.

What is 'at stake'—what may be gained, or is at risk—in selection or assessments during training reflects the significance of the outcomes to those involved. There is likely to be more at risk in summative assessments than in formative assessments, whether these are undertaken during training or as part of selection. Selection to surgical training has been described as 'high-stakes' (Downing, 2003)—in the context of this study, this means that the consequences of selection judgements are prodigious—from individual candidates' perspectives, from the GS training program perspective and from a societal perspective. Candidates consider selection to medical training to be "really high stakes" as "you either get into the career of your choice or you don't" (Kelly, Dowell, Husbands, Newell, O'Flynn & Kropmans et al., 2014, p. 6). From a broader viewpoint, the consequences are similarly high: surgeons are trusted to safely and competently perform operations, their actions affect people's ongoing health and lives. Dijkstra, Galbraith, Hodges, McAvoy, McCrorie & Southgate et al. (2012) suggest that the higher the stakes, the more robust the supporting information needs to be.

Identification of desired attributes of surgeons

Cognitive ability, visuospatial skills, psychomotor skills and personality traits contribute to surgical proficiency. The Royal College of Physicians and Surgeons of Canada (RCPSC) were among the first to encapsulate and specify the range of skills required of clinicians when they developed the seven CanMEDS roles in 1996 (see Figure 4), ("CanMEDS // About", 2017). These roles, regularly updated since inception, continue to describe the abilities physicians draw on. The CanMEDS framework places 'Medical Expert' as the key role of medical practitioners, reinforcing the six other roles, but also being augmented by them.

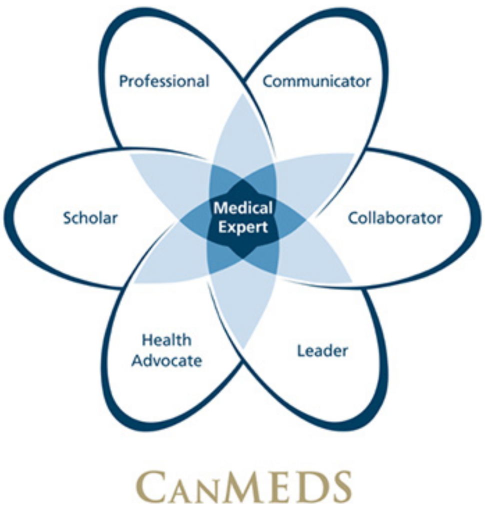

Figure 4. CanMEDS framework

Reprinted from *CanMEDS interactive*, 2017, Retrieved from http://canmeds.royalcollege.ca/en/about. Copyright 2015 by The Royal College of Physicians and Surgeons of Canada.

In 2003 the CanMEDS roles were modified by RACS to define nine competencies that were more specifically applicable to surgeons (see Figure 5) (*"Competencies | Royal Australasian College of Surgeons", 2017*; Watters & Civil, 2011). These nine RACS competencies were similar to the CanMEDS roles, but focussed on skills rather than roles and, reflecting procedural aspects of surgery, included 'Technical expertise' and 'Judgement and clinical decision–making' in addition to the seven competencies readily identifiable with the CanMEDS roles.

Collaboration and teamwork	Ability to work co-operatively with peers, trainees and other health professionals to develop a shared picture of the clinical situation and facilitate appropriate task delegation, to ensure the safe delivery of safe, effective and efficient surgery.
Communication	Communicating effectively with patients, families, carers, colleagues and others involved in health services to facilitate the provision of high quality health care.
Health advocacy	Identifying and responding to the health needs and expectations of individual patients, families, carers and communities.
Judgement and clinical decision–making	Making informed and timely decisions regarding assessment, diagnosis, surgical management, follow-up, health maintenance and promotion.
Management and leadership	Providing direction, promoting high standards, matching resources to demand for services and showing consideration for all members of staff.
Medical expertise	Integrating and applying surgical knowledge, clinical skills and professional attitudes in the provision of patient care.
Professionalism	Demonstrating commitment to patients, community and the profession through the ethical practice of surgery.
Scholarship and teaching	Demonstrate a life-long commitment to reflective learning, and the creation, dissemination, application and translation of the medical knowledge
Technical expertise	Safely and effectively performing appropriate surgical procedures.

Figure 5. Nine RACS competencies
(“Competencies | Royal Australasian College of Surgeons”, 2017; Watters & Civil, 2011)

Notwithstanding these sets of competencies, others have also grappled with defining what epitomises a proficient surgeon. Schueneman, Pickleman, Hesslein and Freeark (1984) recognised over 30 years ago that non-cognitive attributes were as important as surgical dexterity: “Contrary to surgical folklore, pure psychomotor skill is not the major dimension distinguishing between the proficient surgical performance from the mediocre. Rather, non-verbal, visuospatial problem solving abilities appear most crucial to superior technique” (p. 293). While conceptually there is increasing agreement that non-cognitive skills are important, defining specific attributes and their relative importance has become grist for many studies. Sharma (2015) refers to “the five professional attributes: commitment to professionalism, coping with pressure, effective communication, patient focus and working effectively as part of a team” (p. 237). Tansley, Kakar, Withey and Butler (2007), comparing surgical with aviation training, emphasise dexterity, visuospatial and technical ability, while Cuschieri et al. (2001), from their Delphi study involving 44 surgeons, proposed a more comprehensive fusion of technical (operative) and clinical skills and up-to-date medical knowledge, with “personality traits” (p. 110) enabling good judgement, commitment, and “genuine concern” (p. 110) for patients’ welfare. Participants described desirable personality traits as: decision making, integrity, emotional stability, empathy, work ethic, ability to cope with stress, adaptability, and organisational ability. Hall, Ellis and Hamdorf (2003) propose leadership as a major characteristic of effective surgeons, with decision–making and drawing on “higher–order cognitive skills” (p. 14), being considerably more important than manual skill when conducting operations. Foster, Neidert, Brubaker-Rimmer, Artalejo and Caruso (2010) identified surgeons’ preference for public service (helping others), the sciences (empirical research), business relations (basing business decisions on data) and managerial (directing others’ work). Adam, Bore, McKendree, Munro and Powis (2012) refer to doctors’ academic ability and personal qualities of teamwork and professional skills, duty and responsibility, professionalism and values, communication and interpersonal skills, trustworthiness and ethical behaviour, while others (Arora, Sevdalis, Nestel, Woloshynowych, Darzi & Kneebone, 2010; Beasley, 2015; Hall et al., 2003) add the abilities to make decisions quickly, to accommodate uncertainty, ambiguity and incomplete or contradictory information, to contend with serious consequences and to manage stress. The British General Medical Council (GMC) (*Workplace based assessment: A guide for implementation*, 2010) enumerates additional ethical and psychosocial considerations and humanistic qualities, including duty, honour, humility and cultural sensitivity. The physical and cognitive attributes of endurance and the ability to maintain focus on "intricate technical details" are considered essential by Anton, Montero, Howley, Brown and Stefanidis (2015, p. 846).

Bann and Darzi (2005) highlight the uncertainties of trying to encapsulate qualities that define a surgeon when they recognise that "no one particular aptitude or personality study can conclusively provide evidence of the superior visual–spatial ability of a surgeon or the need for a particular attribute to be a successful surgeon" (Bann & Darzi, 2005, p. 101). This is consistent with Wanzel et al's (2002) view that "there is certainly not a single prototype of a good surgeon" (Wanzel et al., 2002, p. 596).

Surgery is a challenging and rewarding career requiring commitment, discipline and compassion (*Workplace based assessment: A guide for implementation*, 2010; Wanzel et al., 2002). The profession presents numerous challenges in its practice and in training practitioners; the consequences of errors, or unforeseen, or chance events may be life threatening. Surgeons practise in complex, stressful, "ever-changing" environments (Foster et al., 2010, p. 366); they prioritise multiple tasks, solve intricate problems and make ethically- and morally-laden as well as medical decisions, often with rapidly changing or inadequate information. Surgeons resolve uncertainty, manage high cognitive loads, and must act decisively and quickly when necessary (Sharma, 2015; Wanzel et al., 2002). Surgeons work with others—they therefore must communicate appropriately, work effectively within local health systems and in professional teams and must reflect to determine when to seek help; surgeons also interact with complex apparatus and instruments. Throughout, surgeons must maintain focus on the patient (Foster et al., 2010; Sharma, 2015). The list of knowledge, skills and attributes required by surgeons seems inexhaustible and ever–changing. Schaverien (2016) and Anton et al. (2015) encapsulate this when they refer to “continually emerging complexities in the field of surgery” (Schaverien, 2016, p. 725) and “increasingly difficult and constantly evolving surgical procedures” (Anton et al., 2015, p. 846).

Identification of desired attributes of trainees

Researchers have grappled with defining desired attributes of surgical trainees and with using these to inform selection criteria. Many authors, including Burgess et al. (2014), advise that recruiting and selecting the best candidates is central to the success of postgraduate training programs and the quality of the medical workforce. But who are the ‘best’ candidates? What attributes are sought in candidates and trainees? Cuschieri et al.’s (2001) study of attributes considered important in the selection of surgical trainees suggested that the attributes that would benefit trainees during training, and were thus important selection criteria, comprised cognitive factors, innate dexterity and personality traits. The researchers expressed “innate dexterity” as spatial perception, hand–eye coordination, aiming, multi–limb coordination and hand–arm steadiness. The Macquarie Dictionary (Delbridge, Bernard, Blair, Butler, Peters & Yallop, 2003) defines cognition as “the act or process of knowing; perception” and cognitive processes as “the ability to acquire knowledge by the use of reasoning, intuition or perception” (p. 380). Cognition and cognitive processes have been linked with success in training (Farkas, Nagpal, Curras, Shah & Cosgrove, 2012; Hall et al., 2003). Hall et al. (2003) consider cognitive skills to relate more to decision–making, proposing that “higher–order cognitive skills involve the ability to identify central issues and assumptions in an argument, recognize important relationships, evaluate conclusions drawn from data, and appraise evidence” (p. 14). Thus, cognitive facility reflects not only the ability to acquire knowledge, but perception, reasoning and decision–making abilities. Some researchers additionally propose links between cognitive competence and clinical performance.

Van de Loo (1988) presents five alternative criteria for selection of surgical residents: “intelligence (verbal, spatial, numeric), operative skill (dexterity, psychomotor ability, attention, concentration), stability and organisation (stress tolerance, judgement, organisational ability), work attitude (motivation, accuracy and carefulness, energy), and cooperation (sociability, independence, self–criticism, empathy)" (p. 278). Bell et al. (2011) contemplated "intangible characteristics" (p. 534) in a survey of 75 applicants to surgical training. Factors predicting success in their program included "an independent desire for knowledge, a commitment to the service of others, and a … sense of direction and purpose" (Bell et al., 2011, p. 534). Gardner et al., (2016, p. 534) consider trainees' responsiveness to the workplace, contending that success in training relies on an ability to "train, adapt, and perform in highly stressful and dynamic situations". Gardner et al also suggest that the skills required to perform well may differ at various stages throughout a training program.

Difficulties in identifying the desired attributes of trainees are similar to those encountered when attempting to define surgeons' optimal attributes. Carmichael et al. (2005) affirm "the definition of a successful resident is not clear" (p. 532) and, while most programs agree "in the abstract" (p. 532) on the important attributes of a good resident (Gilbart, Cusimano & Regehr, 2001), how they are described, and each program's perception of how they value these dimensions, differs. Such differences presumably reflect unique training contexts, curricula and educational philosophies (Kulatunga-Moruzi & Norman, 2002). The extent and diversity of opinions about desirable qualities of both surgeons and trainees make it difficult to synthesise a response beyond the superficial. While differing, context–specific priorities influence particular attributes sought in trainees, in general, most are likely to agree that cognitive factors, physical skills (psychomotor) and personality traits (affective) all contribute to trainee performance (Dirschl, Dahners, Adams, Crouch & Wilson, 2002). The specifics of attributes sought and testing methods vary per program.

Current skills or potential? Aptitude and learning

Compounding selection considerations is the question of whether selectors are seeking evidence of candidates' established, current skills, or indications of yet undeveloped skills—candidates' potential—or a combination of these. Affiliated with this is the question of whether all relevant skills can be learned—if skills can readily be learned, then it may suffice to select those with potential, if skills cannot be learned, then evidence of candidates' inherent proficiency at a stated level may be required.

Kulatunga-Moruzi and Norman (2002) speculate that characteristics such as integrity, leadership, communication, work ethic and orientation toward service "are not easily taught and therefore must be included in the selection process" (p. 35). However, they also deem that assessing these qualities is fraught with difficulty—simultaneously expanding the range and complexity of selection and challenging the validity and reliability of assessment instruments in this field. A counter position is posed by Cruess, Cruess and Steinert (2009) who are not only adamant that professionalism can be taught and learned, but that it "must be taught explicitly" (p. 74) throughout the continuum of medical training and continuing professional development. Cruess et al. (2009) propose initially teaching the "cognitive base" of professionalism—an activity that they maintain is "not difficult" (p. 74)—and subsequently introducing discussion, reflection and assessment to develop "professional identity" (p. 79). Cruess et al. (2009) discuss learner receptivity and ability, and propose opportunely timed instruction and activities of graduated difficulty. This is commensurate with starting instruction in the cognitive range (knowing and understanding) of Miller's pyramid (Miller, 1990) and progressing to the increasing complexities of 'doing' and 'being'. Others (Arnold, 2002; Baldwin, 2003) consider whether characteristics associated with professionalism may follow "their own developmental sequences" (Baldwin, 2003, p. 8) throughout stages of medical training and careers—implying a less explicit, more innate, evolutionary approach to development and consolidation of professional skills and attributes. Tansley et al. (2007), as mentioned previously, emphasise the importance of visuospatial and technical skills, proposing that only candidates with innate ability in these skills should be selected for training.

The concept of latent ability may have been over-emphasised by some. The admonition: "born to be good, train to be great" (Hall et al., 2003, p. 12) implies that innate ability will only take one so far, but to achieve excellence, training and practice are required. Some researchers (Bishawi & Pryor, 2014) maintain that the learning curve—the rate of skill acquisition—varies greatly, dependent on trainees' "baseline" aptitude (Buckley, Kavanagh, Nugent, Ryan, Traynor & Neary, 2014, p. 269). They contend that high aptitude leads to a faster learning curve and improved performance, while low "innate ability" may result in lower proficiency with a "fixed upper plateau" despite practice (Van Bruwaene, Lissens, De Win, Neyrinck, Lens, Schijven & Miserez, 2015, p. 1247). Ostensibly, those with greatest aptitude will progress most efficiently through training and will achieve higher standards of performance. However, Ericsson (2015) counters this, citing "deliberate practice" (p. 1472)—individualised training, targetted at specific aspects of performance, informed by feedback and characterised by repetition and "effortful" concentration—as being fundamental to the acquisition and maintenance of skills. In Ericsson's (2015) "expert–performance" training model (p. 1476), a goal of specific expert performance is identified and studied for its "mediating characteristics" (p. 1476). These characteristics provide the basis for cognitive training and performance feedback, guiding learners' 'deliberate practice' to integrate the new skills with their existing knowledge and skills. Ericsson (2015) thus emphasises the importance of practice and ongoing experience to the development of expertise, a position consistent with Wanzel, Hamstra, Caminiti, Anastakis, Grober, and Reznick (2003) who found that practice and experience appear to obviate the impact of "innate abilities" (p. 757).

This stance does not negate the notion that some individuals will perform at higher standards than others; many considerations influence performance throughout trainees' learning trajectories and beyond. A consideration therefore, is for training programs to define and assess minimum standards of proficiency at selection, and to employ training strategies to facilitate skill acquisition. This strategy de-emphasises the value of searching for 'innate' skills or 'potential' at selection, or seeking those who are already skilled (Gardner et al., 2016).

Cognitive, visuospatial and psychomotor skills in selection

In selection, importance is often placed on discerning candidates' cognitive abilities and in their performance across a spectrum of technical and non-technical competencies. Relationships between cognitive assessments and later performance are inconclusive. Although some say that cognitive tests do not predict achievements during training nor physicians' performance (White, 2002), others have identified cognitive assessments (undergraduate GPA, Medical College Admissions Test (MCAT) verbal reasoning scores and UK A level grades) to predict not just academic but also clinical performance (Dirschl et al., 2002; Kulatunga-Moruzi & Norman, 2002; McManus, 2003), a proposition that implies that cognitive and non-cognitive qualities are not mutually exclusive (Eva & Reiter, 2004). Academic achievement has also been found to predict successful completion of surgical training (Maan et al., 2012). Cognitive ability is not a synonym for 'intelligence'. General intelligence scores can include test results for language/verbal, reasoning, memory, spatial and psychomotor abilities (Deary, 1998). It has been argued that intelligence shows stability through life (McManus, 2003), although this is not universally agreed (Deary, 1998; Forsythe & Johnson, 2016).

Despite traditionally being the dominant selection criteria, academic success and cognitive skills are now considered to be inadequate measures by which to select surgical trainees as they are unlikely to meet the breadth of current requirements. Selection models that rank candidates by their combined academic achievements and their performance in a traditional interview have also been criticised (Grantcharov & Reznick, 2009). However, most surgical training programs in Europe, Canada and North America still select trainees based primarily on "cognitive factors … and academic achievements" (Cuschieri et al., 2001, p. 111; Gallagher et al., 2009; Vassiliou & Feldman, 2011). Most researchers affirm that selection of surgical trainees must delve beyond academic achievements. There is less agreement, however, regarding other attributes or skills to be tested, or on the means to do so. Eva and Reiter, (2004, p. 166) propose assessing non-cognitive qualities in selection. Gallagher et al. (2009) propose that as "technical skills and fundamental abilities (psychomotor skills, visuospatial ability and depth perception)" (p. 109) are critically important to surgical practice, they therefore should be assessed in selection. Cuschieri et al. (2001) found no evidence to support the use of cognitive, psychological, or psychomotor paper tests in screening candidates—advising that, cognitive tests (i.e. examinations) "were considered inappropriate at this stage" (p. 112), and psychological tests "lack validation in recruitment to the medical profession" (p. 112).

Visuospatial (or visual–spatial) ability refers to the visual processing of spatial relations of image properties. A hierarchy of visuospatial abilities has been identified, from simple identification and differentiation of shapes and edges, to complex "whole-object processing" (Anastakis, Hamstra & Matsumoto, 2000, p. 470; Maan et al., 2012; Wanzel et al., 2003). In the surgical context, visuospatial ability describes surgeons' capacity to perceive, analyse and estimate relationships of objects in space (White, 2002)—to mentally manipulate two- and three-dimensional figures (Buckley et al., 2014). All surgical tasks require low level visual processing and some tasks involve high level, spatially complex visualisation of anticipated outcomes (Anastakis et al., 2000). Visuospatial perception has been found to predict rate of skill acquisition and quality of surgical performance (Maan et al., 2012), although Wanzel et al. (2003) imply, and Tang, Hilsinger, Cruz Schloegel, Byl, and Rasgon (2014) conclude that "practice and experience can compensate for initial discrepancies in visuospatial ability and manual dexterity" (Tang et al., 2014, p. 248). These researchers propose that visuospatial ability appears to be more crucial initially and its importance decreases as practitioners gain experience performing a procedure.

Psychomotor skill refers to dexterity, or "the ability to perform motor tasks with precision and coordination" (Buckley et al., 2014, p. 265). Psychomotor aptitude has been shown to correlate with visual–spatial aptitude (Van Bruwaene et al., 2015), with rate of skill acquisition (Maan et al., 2012) and with resident performance (Dirschl et al., 2002). Tests of manual dexterity aptitude in selection can be controversial. Some tests have been criticised for subjective ratings and poor definition of 'surgical skill' (Tang et al., 2014). Tests for motor skills as well as outcomes, differ. While Dirschl et al. (2002) found motor performance to correlate with overall resident performance, a later study found that performance in soap carving tests in selection to otolaryngology training did not correlate with later clinical performance (Tang et al., 2014).

Although trainee behaviour is impossible to predict accurately, selectors still try to identify candidates' attributes and use these as surrogates for potential (White, 2002) or aptitude. Selection to surgical training usually relies on identifying and placing values on particular achievements and abilities, using the outcomes of assessments for these achievements and abilities to rank candidates and to admit those who meet minimum criteria. Tests of achievement usually rely on "recent [formal or informal] educational experience" (McManus, 2003, p. 139)—they convey candidates' performance at the time of the assessment and may be used to establish that candidates have attained core knowledge or skills. Tests for cognitive, visuospatial and psychomotor skills have all been used as selection criteria, however their importance and accuracy in determining performance during training have not been resolved.

Many studies (for example, Dirschl et al., 2002; Eva Reiter, Trinh, Wasi, Rosenfeld & Norman, 2009; Geissler, VanHeest, Tatman & Gioe, 2013; Lievens & Coetsier, 2002; McManus, 2003; Maan et al., 2012; Neely, Feinglass & Wallace, 2010; Patterson, Lievens, Kerrin, Munro & Irish, 2013; Schueneman et al., 1984), including the current one, seek to provide evidence regarding the predictive power of selection—the extent to which candidates' current performance may indicate future performance. Studies strive to identify the domains in which predictive relationships are most likely to occur and the selection instruments that best signal future performance. The outcomes of such studies inform future practice.

In selection, selectors are looking for indications of specific knowledge and skills being present to meet predetermined requirements. Tests of achievement identify current or past knowledge and skills. Assessments in selection are also implemented to anticipate candidates' likely performance during training and beyond, although such latent ability is near–impossible to predict accurately. There is, therefore, a divergence in the focus of these assessments—they measure current performance, reflect past experiences and anticipate likely future performance.

Tests of knowledge are comparatively unambiguous, being used to authenticate candidates' core knowledge of subjects deemed relevant to training, (e.g. anatomy, pathology, physiology), thus establishing a cognitive foundation from which to develop subsequent activities. In selection, such tests may be prerequisites—hurdles that candidates must achieve—or they may be scored selection components, contributing to candidates' rankings. Some tests of skills, notably, tests of tangible, concrete skills, are similarly straightforward, but other tests may be more subjective or imprecise—particularly tests of non-cognitive attributes, latent abilities and those pitched at the more complex levels of Bloom's taxonomy (Bloom, 1956) and Miller's pyramid (Miller, 1990). Inexpert preparation of questions or suboptimal assessments of responses are more likely to take place and to impact on outcomes in non-cognitive domains.

Some tests for 'intangible characteristics' or personality traits, assume that candidate performance represents latent or innate ability. Many studies have explored the predictive capacity of selection instruments. In contrast, situational judgement tests (SJTs) (discussed more fully later in the Literature Review) assess candidates' actual skills in judgement and decision–making. SJTs have been successfully used in other professions and have been introduced in selection to some medical training programs. As is the case with many selection instruments, the quality of questions impacts on the reliability of the assessment protocol.

Diversity

Recent pressures to promote diversification of the medical workforce are an acknowledgement that this workforce could better serve the community by being more representative of it. Recognition that benefits ensue from increased diversity has permeated many facets of recruitment and employment (Armstrong, 2006), although appreciation that the medical workforce and the society it serves stand to benefit from increased diversity (Cleland, Patterson, Dowell & Nicholson, 2014) is a relatively recent phenomenon. Valuing diversity in the workplace means valuing the differences between people and the different qualities they bring to their roles and recognising the likelihood that these differences can lead to more rewarding work environments (Armstrong, 2006).

Australia's "changing and increasingly diverse society" (Scanlon Foundation, 2016, p. 1) has become one of its most defining characteristics. In 2015 people born overseas comprised more than a quarter (28.2%) of Australia's population (3412.0 - Migration, Australia, 2014-15, 2016); New Zealand's population is similarly becoming more ethnically diverse (2013 Census – Major ethnic groups in New Zealand, 2015; National Ethnic Population Projections: 2013–2038, 2017). Fortunately, most Australians feel that multiculturalism is good for the country (Scanlon Foundation, 2016). In addition to culturally and linguistically diverse backgrounds however, these populations face issues related to age, gender, sexuality, disability, legal status, faith and many other factors (Deptartment of Health and Human Services, 2016). One local response, produced by the Victorian Department of Health and Human Services, is a cultural diversity plan highlighting the varied health characteristics and requirements for care of the multicultural population it serves. The plan recognises that the department's responsiveness can be "greatly enriched" (p. 17) by a similarly culturally diverse health workforce who are supported to "improve their cultural competence" (p. 17). Notwithstanding the importance of ethnicity, cultural diversity also encompasses people's gender, sexuality, socio-economic status, religious beliefs, geographic origins, education and other influences (Razack, Hodges, Steinert & Maguire, 2014).

The UK Chartered Institute of Personnel and Development (CIPD) proposes that organisations develop "inclusive approaches to employment…practices" in order to promote diversity in the workforce, advising that valuing diversity goes beyond anti-discrimination and 'equal opportunity' (CIPD, 2015). Considering such an approach when recruiting medical students, Cleland and colleagues, in more than one study (2012, 2014) advocate "widening access" (2014, p. 20) to medical education to ensure that doctors are representative of the society in which they will practise. They advise that students who are exposed to diversity in training may "gain a greater understanding of … people from different socio–cultural backgrounds," (2012, p. 50) and thereby gain an increased ability to provide healthcare to people from different backgrounds. White (2002) notes that diversity among health education participants can bring different perspectives to enrich analysis, perception and interpretation of knowledge, while Kwakwa and Jonasson (1999) propose that certain population groups are best served "by physicians who are aware of and sensitive to cultural and language differences" (p. 584).

Selection methods may unintentionally exclude those from disadvantaged and low socio-economic backgrounds, some ethnic and cultural groups, rural and regional applicants and those with disabilities (Cleland et al., 2012; Girotti, Park & Tekian, 2014; Prideaux et al., 2011). Competitive selection protocols may penalise candidates who lack access to developmental opportunities, disadvantaging them when competing against individuals from more favourable backgrounds. Students from disadvantaged backgrounds may also encounter associated challenges during training, warranting supplementary measures to support their progress (Girotti et al., 2014).

Issues to be addressed in building a more diverse medical workforce include attracting a more diverse applicant pool, ensuring that selection methods do not inadvertently discriminate against people from 'under-represented' groups, considering affirmative actions to compensate for disadvantage and implementing programs to support students from disadvantaged groups to succeed in training. As Lievens (2014) summarises, these strategies may be categorised as "attraction–selection–inclusion–retention" (p. 11). However, he also challenges that selection processes that "admit a diverse student pool" (p. 12) may have reduced validity—the "validity–diversity dilemma" (p. 12). People from minority groups have been shown to score lower on "the most valid tests" (p. 12) possibly because they typically have less test familiarity and fewer test–taking skills. However, Hay, Mercer, Lichtwark, Tran, Hodgson and Aretz et al., (2017) counter that reducing selection scores "does not result in increased failure, or impaired performance during [medical] training" (p. 546).

Individuals with unique attributes and from diverse backgrounds can be successful in surgical training. Equitable selection processes will evaluate and balance contrasting components and influences; they will maintain standards and facilitate diversity, recognising that production of identical clones is not desirable, even if it were achievable (Bell et al., 2011; Glick, 2000). To increase diversity in the medical workforce, it may be necessary to address inequities by compromising on some of the selection considerations—illustrating Lievens' (2014) "validity–diversity dilemma" (p. 12) and to implement additional supports during training. This study recognises that diversity is an area that RACS may choose to consider when reviewing selection.

Key issues in surgical training and selection – summary

This section has explored some of the key attributes sought in candidates aspiring to surgical training. There is limited agreement about their relative importance and how to address them. There has been an historical reliance on testing applicants' cognitive and academic ability although there is no consensus on the relevance of cognitive assessments in selection. Cognitive and non-cognitive qualities are not mutually exclusive. The current consensus is that selection of surgical trainees must delve beyond academic achievements. Technical skills, psychomotor skills, visuospatial ability and depth perception can be assessed in selection, although some tests have been criticised for their subjective ratings and a poor definition of surgical skill. There is debate regarding whether selectors are assessing candidates' current abilities or their potential for future performance. However, at present, predictive relationships can only be determined retrospectively, by studies such as this one. There is evidence that skills can be taught and improved with practice and experience.

The laudable aim of increasing the diversity of surgical trainee cohorts may compromise the validity of selection assessments, (invoking at-times divergent processes for attraction, selection, inclusion, and retention). To increase diversity in the surgical workforce, it may be necessary to modify some selection protocols.

The next section reviews some key concerns regarding the selection instruments and processes. These include discussions of validity and reliability, feasibility, acceptability, bias and fairness.

Part 3 Attributes of selection and assessment instruments and processes

Validity

In the context of this study, the validity of a selection or assessment instrument refers to the extent to which it measures what it was intended to measure (van der Vleuten, 2000). Several forms of validity have been identified, those particularly pertinent to this study being:

Face (or content) validity: the item or content 'makes sense', or appears relevant to those involved (Cleland et al., 2012; Koczwara, Patterson, Zibarras, Kerrin, Irish & Wilkinson, 2012; *Principles for the Validation and use of Personnel Selection Procedures*, 2003; Beard, Marriott, Purdie & Crossley, 2011; Michels, Avonts, Peeraer, Ulenaers, Van Gaal, & Bossaert, et al., 2016; Twycross & Shields, 2004)—in this study, this would refer to the item's perceived relevance to surgical training.

Criterion validity: the candidates' scores (outcomes) in one or more criteria correlate with an external reference or outcome. Such performance comparisons may be with parallel assessments in related areas (concurrent validity) or with performance in subsequent assessments (predictive validity) (Beard et al., 2011; Crossley, Humphris & Jolly, 2002; Downing, 2003; Newton & Shaw, 2015; Twycross & Shields, 2004). This study reviews predictive validity of performance in three selection instruments.

Construct validity: the extent to which a test measures the 'construct' that it is intended to measure and can differentiate between candidates or groups who differ in ability (Beard et al., 2011; Cleland et al., 2012; van der Vleuten, 2000;). Some consider that construct validity is the prime form of validity (Downing, 2003; Fromme, Karani & Downing, 2009).

Incremental validity occurs when combining the scores of assessment items increases their predictive validity above that of any of the individual assessments (Lievens, Peeters, & Schollaert, 2008)—in this case, when performance in two or more selection instruments shows greater predictive validity than does performance in any of the instruments individually.

A major evaluation standard by which to judge the effectiveness of selection procedures is their capacity for predictive validity (Patterson, Baron, Carr, Plint and Lane, 2009). Many studies have been conducted to ascertain relationships between selection criteria and subsequent performance across the breadth of medical specialties and throughout the continuum of medical training. Predictive validity is usually expressed as a correlation coefficient (Andriole, Jeffe & Whelan, 2004; Dirschl et al., 2002; Dowell, Lumsden, Powis, Munro, Bore, Makubate & Kumwenda, 2011; Gallagher et al., 2009; Maan et al., 2012; Prideaux et al., 2011; Schaverien, 2016; Schueneman et al., 1984). A correlation of one indicates a 'perfect' association while a correlation of zero signifies that the 'predictor' has no predictive relationship with the criterion (Wiliam, 2001).

As Downing (2003) states, "assessments are not valid or invalid; rather, the scores or outcomes of assessment have more or less evidence to support (or refute) a specific interpretation" (p. 830). He proposes that key evidence supporting selection activities is "the predictive relationship" between selection scores and later achievements—their predictive validity.

Predictive validity has also been linked to the 'job relatedness' or face validity of evidence gathered which, in turn, relies on accurate determinations of the work tasks and the knowledge, skills and attributes required to perform them (Maan et al., 2012). However, although synergies between selection and work activities are likely to enhance the 'acceptability' of selection procedures for participants, face validity alone is insufficient in such high-stakes assessments as selection to surgical training. Face validity is a rather subjective type of validation (Gallagher et al., 2003) and is only one of many factors to be considered when designing selection procedures (*Principles for the Validation and use of Personnel Selection Procedures*, 2003). Downing (2003) considers that construct validity is the most important form of validity, embracing "the whole of validity" (p. 831), since assessments deal with 'constructs'—such as educational achievement or ability—that are inferred from candidates' performance. Recommendations by Downing support carefully defining the constructs being tested and maximising the validity of findings by gathering evidence from multiple sources, particularly for high-stakes assessments. This approach to selection is reinforced by researchers such as Edwards, Friedman and Pearce (2013), who found that "the use of multiple tools does result in an increase in predictive validity" (p. 6) and Patterson et al., (2009), who found that "the most accurate overall prediction was obtained using the … SJT in combination with the other measures" (p. 55). Edwards et al., (2013) similarly found that multiple tools increased predictive validity in selection, with each additional tool incrementally adding to predictive validity.

Reliability

In the context of this study, the reliability of a selection or assessment instrument refers to the extent to which its performance is 'dependable'—consistently yielding the same results, when used repeatedly, under similar conditions, with different candidates and assessors (van der Vleuten, 2000; Gallagher et al., 2003). Reliability is a characteristic of the outcome of an assessment, not of the measuring instrument itself (Downing, 2004). Several forms of reliability have been identified, those particularly pertinent to this study being:

Internal consistency/Inter-item reliability: the extent to which the item/s in the in the instrument provide consistent information—the consistency of scores for components of a single assessment question (Dore, Kreuger, Ladhani, Rolfson, Kurtz & Kulasegaram et al., 2010; Murphy, Bruce, Mercer & Eva, 2008).

Inter-station or inter-question reliability: the consistency of scores between assessment stations; the generalisability of a candidate's scores from one assessment station to another (Dore et al., 2010).

Inter-rater reliability: the accuracy, consistency and level of agreement of raters' scores (Downing, 2004)—particularly the extent to which one rater's scores predict another rater's scores (Murphy et al., 2008).

Test-retest reliability: the reproducibility of test scores across subsequent deliveries (Downing, 2004).

Reliability is often expressed as a reliability coefficient expressing the ratio of "true score variance" to "total score variance" (Downing, 2004, p. 1007) indicating the extent to which error in candidates' scores may be due to low reliability of one or more of the above components. The closer the reliability coefficient is to one, the greater the reliability of the construct being measured (Downing, 2004; Murphy et al., 2008). High-stakes assessments require reliability in the order of 0.90, while for moderate–stakes assessments reliability coefficients of 0.80–0.89 may suffice (Downing, 2004; Wiliam, 2001). It is worth clarifying here the difference between correlation coefficients (measuring predictive validity) and reliability coefficients; Gallagher et al. (2003) explain that correlation coefficients are "no more than measures of association, whereas reliability coefficients are measures of agreement … [implying] a sameness or equal value" (p. 1527).

It is common to consider the reliability of items and questions; however, the notion that assessments of clinical performance rely on inter-rater reliability as the main measure of consistency (Downing, 2004, p. 1012; Williams, Verhulst, Colliver, Sanfey, Chen & Dunnington, 2012) is of particular interest to this study. Referee reports, interviews, clinical examinations, DOPS, MiniCEX and end of term assessments all rely on assessments of clinical performance. Implementing practices such as training the assessors, using clear rating guidelines, using multiple assessors, multiple assessment tools in a variety of settings enhance the reliability and validity of assessments (Davis & Ponnamperuma, 2005; Salvatori, 2001).

The relationship between validity and reliability

Reliability and validity can be affected by similar constructs and are inter-dependent; some consider that validity is more important than reliability (Davis & Ponnamperuma, 2005; Wiliam, 2001). The inter-relationships between reliability and validity can be circuitous, affecting assessment efficacy and outcomes. When validity is high but reliability is low, the assessments measure what they were intended to measure but may produce inconsistent results, or results that hinge on who scores the assessment. When reliability is high but validity is low, assessments will produce consistent outcomes, without testing what they were intended to test (Downing, 2004; Wiliam, 2001). Wiliam (2001) reminds us that reliability and validity "are not absolute but degrees" (p. 20) and suggests that "reliability and validity are in tension, with attempts to increase reliability … having a negative effect on validity" (p. 20). However, to be valid an assessment must also be reliable—reliability "sets the upper limit of validity" (Salvatori, 2001, p. 167). Downing (2004) proposes that "if the stakes are extremely high, the reliability must be high in order to defensibly support the validity evidence for the measure" (p. 1009). Ideally, both the validity and reliability of assessments will be high.

Issues affecting validity and reliability

Many factors—associated with the instruments, the processes, the attributes being assessed, the testing environment, the candidates and the assessors—affect the validity and reliability of assessments. Validity is increased when assessors are trained in the objectives of the assessment and the constructs being measured, in using the assessment instruments and in potential threats to validity, such as the influence of bias on judgements (Wiliam, 2001; Salvatori, 2001). Validity may also be increased when assessment instruments incorporate explicit standards and rating guidelines.

Reliability may be increased by improving test items (for example, their clarity and relevance) and the consistency of marking. Wiliam (2001) suggests that refining the scope of the test and making the test longer are also effective measures to increase reliability, but offers the caveat that even large increases to the length of tests may result in relatively small increases in reliability. There is considerable agreement that to increase the validity and reliability of selection or assessment outcomes, multiple tests are required (Dore et al., 2010; Murphy et al., 2008; Williams et al., 2012; Edwards et al., 2013). As Wiliam (2001) emphatically states, "high-stakes decisions should never be based on the results of individual tests" (p. 19) because "even the best tests" (p. 19) can result in atypical scores for individual candidates. Multiple assessments involving multiple assessors provide more comprehensive information. This is consistent with Dijkstra et al.'s (2012) proposal that the higher the stakes, the more robust the information needs to be and the more certainty is required (p. 5).

Feasibility

In the context of this study, the feasibility of a selection or assessment instrument refers to the extent to which its implementation is achievable, within the available resources. Feasibility is a characteristic of the processes and resources required to implement an assessment, not of the assessment itself. Aspects to be considered when analysing an assessment's feasibility may include protocols for implementing the assessment instrument, its cost-effectiveness, the infrastructure and personnel required, the frequency and the time required to perform or analyse the assessment (Davis & Ponnamperuma, 2005; Dore et al., 2010; Murphy et al., 2008; Salvatori, 2001). Feasibility is the extent to which it is practical for the assessment to be implemented—this may include how well it meets the needs of, and is acceptable to candidates and assessors.

Acceptability

The acceptability of a selection or assessment instrument refers to reactions and impressions by those involved in its implementation—as assessors, as candidates, as governance and as administrators. Evaluations of acceptability commonly consider the perceived fairness, face validity, reliability, or feasibility of assessment instruments and processes. Acceptability is a subjective measure, testing alignment between expectations and experiences. Selection processes that are perceived as fair (van den Bos & Miedema, 2000) or as being relevant to a position may be more acceptable to participants (*Principles for the Validation and use of Personnel Selection Procedures*, 2003) Similarly, selection processes in which candidates are satisfied that they can present themselves to their best advantage are likely to be 'acceptable' to candidates—for example, one study of multi-mini-interview (MMI) 'acceptability' found that overall the process was "well received", with most candidates believing that "they could accurately portray themselves" (Dore et al., 2010, p. 61). Another gauge of acceptability may reflect users' 'confidence' in the outcomes of an assessment—Downing (2004, p. 1010) suggests that unreliability of an assessment instrument tends to reduce confidence in assessment outcomes. These subjective evaluations reflect how users perceive assessments, their levels of confidence in the selection or assessment outcomes and may also influence person–environment fit. However, the effect of 'acceptability' on the outcomes of assessments is debatable and would usually be considered to be less important than validity or reliability (*Principles for the Validation and use of Personnel Selection Procedures*, 2003).

Bias

In the context of this study, bias refers to score variations that arise from assessors' idiosyncrasies (Yeates, O'Neill, Mann & Eva, 2012), or from social or environmental influences (Williams, Klamen & McGaghie, 2003). Bias can be explicit or implicit and can favour, or discriminate against, groups or individuals on the basis of attributes such as gender, ethnicity, socio-economic status, cultural, personal or professional characteristics (Cleland et al., 2012; Yeates et al., 2012). Biases can influence all components of the selection and assessment cycle, including question creation, implementation and the interpretation of scores (Tavakol & Dennick, 2011). By introducing score variability, bias can compromise the fairness, validity and reliability of assessments (Yeates et al., 2012). Bias introduced through a particular measurement instrument may be diluted by using multiple instruments; individual scorer biases may similarly be diluted by using multiple tools and sampling widely (Reiter & McConnell, 2014). Although unconscious bias is considered detrimental to fairness, conscious, positive bias can be used to counteract social imbalances. Emery et al., (2011) caution that defining and applying fairness in medical admissions is complex, as fairness and bias are "not synonymous" (p. 63) and that at times bias—in the form of positive discrimination—may be justified, for example to increase diversity in a student population. Examining and addressing potential biases is likely to enhance fairness (Razack et al., 2014).

Fairness

Fairness is a multifaceted concept. It pertains to how people are treated or how they, or others, perceive they are treated, particularly in situations involving allocation of resources (Blake, McAuliffe & Warneken, 2014, p. 559; van den Bos & Miedema, 2000, p. 355)'. Equity is fundamental to fairness—a fair division is commonly considered to comprise equal shares (Cappelen, Nielsen, Tungodden, Tyran & Wengström, 2015). The concept of fairness invokes impartiality, objectivity, honesty and lack of self-interest, prejudice, bias or favoritism. Fairness also draws on notions of 'deservedness'—earning an outcome by demonstrating merit—through diligence, talent, aptitude or other attribute. In a fair process, those who most merit a reward are most likely to receive it. If selection processes are fair, those who are most suited to training are most likely to be selected. Fairness in selection implies that everyone has a fair opportunity and chance of being selected based on talent and merit (The Panel on Fair Access to the Professions, 2009). In selection *processes*, fairness has strong connections to procedural justice, while fairness in selection *outcomes* has been described as "distributive justice" (Patterson, Zibarras, Carr, Irish & Gregory, 2011, p. 290). Recommendations for procedural justice in selection are partly based on the assumption that fair processes and behaviour will lead to fair outcomes.

In their Legal Practice Note: Procedural fairness/natural justice (Australian Health Practitioner Regulation Agency - Legal Practice Notes - LPN 17, 2013), the Australian Health Practitioner Regulation Agency (AHPRA) identify procedural fairness as "the duty cast on administrative decision-makers to act fairly when making decisions which may affect people's rights, interests and legitimate expectations" (p. 1). If, as Liu and Lu (2016) suggest, ongoing interpersonal relationships are affected by resource distribution, then interactions or transactions that are perceived to be fair, or unfair, may affect future exchanges. This resonates with notions of procedural justice as discussed elsewhere in this Literature Review, underscoring the fundamental and intricate nature of fairness in selection.

It is normal to strengthen the fairness of assessments (including selection assessments), by standardising test conditions and test questions so all candidates' experiences are as alike as possible; however, this may only partially address issues of fairness. In their discussions regarding fairness in assessment, Gipps and Stobart, (2009), consider that assessment practices need to be "fair and just for all groups" (p. 106) and that candidates' equity of access to opportunities to prepare for assessments is paramount. If all candidates' learning opportunities and experiences are alike, then assessment outcomes are more likely to reflect candidate ability and therefore be fairer. Conversely, when preparation experiences differ, assessment outcomes may be influenced by factors other than those intended to be assessed. In selection to medical and surgical training, candidates' preparatory experiences may differ greatly, being affected by their social, cultural or economic background, as well as by their gender or ethnicity (Razack, 2016; Robb, Dunkley, Boynton & Greenhalgh, 2007)'. The UK report, *Unleashing Aspiration: The Final Report of the Panel on Fair Access to the Professions* (2009), proposes that fair selection to the professions benefits candidates and employers alike, as it enhances access for underrepresented groups and increases the likelihood that the best candidates are chosen. It can be challenging, though, to satisfy all participants that processes are truly 'fair'.

Cappelen et al., (2015) propose that fair behaviour is intuitive to most people, however, when measures are introduced to compensate some candidates for perceived disadvantage or injustice, notions of fairness can become contentious. When judgements about deservedness intersect with social justice, further complexities and transactional elements to fairness are revealed. Razack (2016) highlights tensions that can exist between theoretical and operational fairness, proposing that activities that are intended to be fair can actually expose implicit values and prejudices and that selection processes are "deeply imbued with social, cultural and economic capital" (p. 601). This indicates that values underpinning selection practices should be explicitly and critically considered when developing selection processes, notwithstanding the "complex terrain with multiple priorities" negotiated by selectors as they attempt to select fairly (Razack et al., 2014, p. 41).

It is challenging to implement processes that perceived as 'fair' by all participants. The transactional nature of fairness may result in some participants feeling disenfranchised.

Candidates expect that selection processes and allocation of training positions will be conducted fairly, according to principles of procedural justice, and that outcomes will be merit based (Razack et al., 2014).

Selection and assessment processes

Medical selection processes usually comprise the implementation of multiple selection instruments. In their operation, these instruments may show evidence of more, or less, validity, reliability, feasibility and acceptability. The model Figure 3 indicates that the outcomes of selection are affected by more than merely the instruments. Patterson et al (2016) discuss the notion of integrating assessments into selection systems, consistent with Schuwirth and van der Vleuten's (2012) approach to programmatic assessment.

Attributes of selection and assessment – summary

Multiple factors impinge on the effectiveness of selection and assessment instruments. The many forms of validity—whether the instruments measure the 'constructs' they were intended to measure, whether they appear relevant to participants, or can differentiate between candidates, whether their use in combination is more effective than in isolation, whether they 'predict' later performance—are all considerations in reviewing the efficacy of selection items. The reliability of selection or assessment instruments—the extent to which their performance is dependable when used repeatedly, under similar conditions, with different candidates and assessors—also has multiple forms. Reliability may be measured. Assessments of clinical performance rely on inter-rater reliability. Reliability and validity can be affected by similar constructs and are inter-dependent. Ideally, both the validity and reliability of assessments will be high.

Validity and reliability may be increased by clearly defining the constructs being tested, by using relevant test items, by using explicit standards and rating guidelines, by training assessors and improving the consistency of marking. A further key factor to increase the validity and reliability of selection or assessment outcomes is to gather evidence from multiple sources, using multiple tests.

Feasibility of implementing assessments is constrained by the resources available, while acceptability is a subjective response, possibly testing the alignment between participants' expectations and experiences. Selection processes that are perceived as fair or relevant may be more acceptable to participants. Bias refers to score variations arising from item developers' or assessors' idiosyncrasies or from social or environmental influences. Conscious, positive bias may be used to counteract social imbalances and increase diversity. Bias may be diluted by using multiple instruments and multiple scorers. Fairness is also subjective, pertaining to how people perceive they are treated. Fairness has strong connections to procedural justice and draws on notions of 'deservedness'. Fairness may be increased by standardising tests, although it is often possible for some participants to feel disenfranchised.

All these selection attributes contribute to the development and implementation of selection and assessment practices. Although in many respects they are compatible, they can also act in divergent ways—for example, increasing a test's reliability may decrease its feasibility. Practitioners must consider the breadth of complex interrelationships and conflicting priorities between factors to develop instruments and protocols to suit local conditions.

The next section briefly reviews some general considerations regarding assessment practices. These include discussions of formative and summative assessment, objectivity and subjectivity and scoring systems.

Formative and summative assessments

The GS SET program implements formative and summative assessments. This approach is consistent with good practice. van der Vleuten, Schuwirth, Scheele, Driessen and Hodges (2010) propose that "all the principles of assessment are interrelated and interacting" (p. 716)—that, as formative and summative assessments serve different purposes, a comprehensive assessment program will include both.

Formative assessment

The main purpose of formative assessments is to provide candidates with feedback on their performance to promote learning (Evgeniou et al., 2013; Nicol & Macfarlane-Dick, 2006). Formative assessments are often described as assessments *for* learning (Schuwirth & van der Vleuten, 2011). They are usually low–stakes; consequently, a "lesser degree" (Beard, 2007, p. 1315) of reliability may be acceptable. Construct validity is important however, as assessors must be able to accurately identify areas for improvement in order to provide relevant feedback—as Holmboe, Huot, Chung, Norcini and Hawkins (2003) observe, "faculty cannot correct errors or deficiencies if they cannot correctly identify the errors and deficiencies" (p. 829). van der Vleuten et al., (2010) recommend inclusion of "qualitatively rich information" (p. 713) such as narrative comments in formative assessments to maximise the effectiveness of feedback. In the current study, the DOPS and MiniCEX are implemented as formative assessments; the ETAs include both formative and summative elements.

Summative assessment

The main purpose of summative assessments is to measure performance, marking candidates' completion of a component or stage in training, justifying promotion to the next level or completion of a course and validating certification (Vassiliou & Feldman, 2011). Summative assessments are often described as assessment *of* learning (Schuwirth & van der Vleuten, 2011). They are usually high-stakes, requiring a high degree of reliability (Beard, 2007). In the current study, the selection instruments (CV, RR and Int), the GenSSE, the SpecSSE and the CE are implemented as summative assessments.

Part 4 Selection instruments and selection protocols – general

Objectivity and subjectivity

It is common to equate objectivity with fairness. Being objective is to be impartial and "free from personal feelings or prejudice; unbiased" (Delbridge et al., 2003, p. 1320)—objective decisions are based on quantifiable, factual data. Subjectivity involves emotions, feelings and partiality— "existing in the mind; belonging to the thinking subject rather than to the object of thought" (Delbridge et al., 2003, p. 1871). 'Objective' judgements are often considered to be more reliable than are subjective judgements (Auewarakul, Downing, Jaturatamrong & Praditsuwan, 2005; Bann, Davis, Moorthy, Munz, Hernandez & Khan et al., 2005; Schaverien, 2016). The most objective assessments tend to be those in which the influence of human judgement is minimised—measures of knowledge, such as written, multiple choice exams (Reid, Kim, Mandel, Smith & Bansal, 2014) and assessments of motor skill in which performance is tracked, for example, by virtual reality simulators (Bann & Darzi, 2005). Assessments of non-technical attributes rely on raters using perceptions to form judgements and are thus more subjective. To minimise subjectivity, controls, such as structure and uniformity, can focus processes and content on attributes being assessed and promote similarity in participants' experiences (Gallagher et al., 2014). When variation in assessment content and experiences are minimised, it is assumed that differences in assessment outcomes reflect differences in candidate performance. In addition, the effects of subjective rater biasses may be moderated by using multiple assessment instruments and sampling widely (Reiter & McConnell, 2014).

Many training programs and regulators link objectivity with validity, reliability and fairness, espousing objective selection and assessment practices as not only desirable, but essential (Commonwealth Department of Health and Family Services, 1998; Gallagher et al., 2014; Shellito, Osland, Helmer & Chang, 2010). However, it is impossible to completely eliminate subjectivity in assessments in which human judgement is a factor; as Reiter and McConnell (2014) recognise: "if there was complete objectivity … all [judgements] would be unanimous. Individual voter bias is not only condoned but celebrated as providing unique and valuable perspectives (p. 1142).

Objective assessments in selection have been shown to predict later performance (Shellito et al., 2010). However, these links tend to be to similar, objective assessments. Makdisi et al.

(2011) found no direct correlation between academic performance markers and success as a resident. Desired attributes of trainees encompass more than traits that lend themselves to objective assessment. It is reasonable to assess trainees across the range of desired attributes using both objective and subjective measures. Similarly, in independent practice, surgeons are evaluated objectively (by audit reports) and subjectively (by patients, colleagues, employers). Despite efforts to maximise objectivity in assessments of non-technical attributes, human judgements are integral to the outcomes.

Scoring systems

Scoring systems vary between criterion–referenced, in which candidates are scored relative to defined performance markers or standards, and norm–referenced, in which candidates are scored relative to each other. These systems may be implemented in selection and in assessments during training.

Selection protocols may include combinations of prerequisite achievements (eligibility criteria must be satisfied before applications may be submitted), hurdle assessments (candidates must fulfill these requirements to progress in the selection process) and cumulatively scored assessments (scores in two or more activities are aggregated). The use of standardised measures may facilitate comparison of applicants (Katsufrakis, Uhler & Jones, 2016).

Selection instruments and protocols – general. Summary

It is common to equate objectivity with validity, reliability and fairness. Objective decisions, based on quantifiable data, tend to be those in which human judgement is minimised. Subjective decisions rely on raters using perceptions to form judgements. To minimise subjectivity, controls, such as structure and uniformity may be implemented. It is reasonable to assess broadly, using both objective and subjective measures. Criterion–referenced and norm–referenced scoring systems are variously used in selection and assessment instruments. Combinations of assessments may be implemented as prerequisites, hurdles, or added to arrive at a cumulative score.

The next section reviews some specific selection instruments and selection protocols. These include AOA membership, assessment centres, ballot, curricula vitae, referee reports, interviews, multiple mini interviews, and situational judgement tests.

Selection instruments and selection protocols – specific

In selection to medical training, the selection instruments (or selection 'tools') are the activities implemented to rank candidates for entry to training. Selection protocols (or selection methods) are the processes for implementing the instruments. This section critically reviews an array of instruments frequently used in selection to specialty training and includes some instruments more often used in selection to generic medical training, where this implementation might be analogous to specialty training. Selection instruments are presented in alphabetical order.

Alpha Omega Alpha honor society (AOA) membership

Academic achievement has long been used as a method of determining entry into medical and specialty training. In the USA membership of Alpha Omega Alpha Honor Medical Society (AOA) is offered to medical students who have excelled academically, demonstrated professionalism, "a firm sense of ethics" (Singletary, 2010, p. 368) and "shown promise of becoming leaders in the profession" (Maan et al., 2012, p. 1616). A 2007 study revealed that approximately 20% of students entering USA surgical residency programs were AOA members (Bell et al., 2007).

AOA membership is considered to be an objective selection tool—candidates for specialty training are either members or not—and many USA program directors rate AOA membership highly when evaluating candidates. Fifteen Plastic Surgery program directors surveyed by LaGrasso, Kennedy, Hoehn, Ashruf and Przybyla (2008) considered AOA membership to be "the most important objective criterion" (p. 122e) used to evaluate candidates while, in another study, General Surgery program directors rated AOA membership as a "primary factor" in residency selection (Melendez, Xu, Sexton, Shapiro & Mohan, 2008, p.

151). However, as Cone et al. (2012) observe, “AOA membership is awarded according to different criteria at different institutions” (p. 263) and is therefore less objective than may have initially been surmised.

Studies, such as those reported by Andriole et al., (2004), Carmichael et al., (2005) and Cullen, Reed, Halvorsen, Wittich, Kreuziger and Keddis, et al., (2011), have found that AOA membership does not significantly predict performance in specialty training, nor do AOA members perform significantly better than non-members, although Carmichael et al. (2005) found 60 Orthopaedic surgical trainees who were AOA members “had slightly higher average Orthopaedic In-Training Examination (OITE) scores than those who were not members” (p. 532). A single study—conducted by Kron, Kaiser, Nolan, Rudolf, Muller and Jones (1985)—was cited in Maan et al.’s review of 24 projects (2012) as reporting a significant positive correlation between AOA membership and “successful completion of training” (p. 1616).

In summary, AOA membership is recognised as a reliable measure of academic achievement, is simple to implement as a selection criterion and is attractive to many USA program directors; AOA membership has also been associated with completion of training. However, AOA membership may not be as objective as it is popularly considered to be nor has it been shown to predict performance during training.

Assessment centres/Selection centres

Assessment centres—also known as selection centres—implement multiple standardised assessment exercises to provide opportunities for multiple raters to evaluate applicant behaviors (Armstrong, 2006; Koczwara et al., 2012; Monroe, Quinn, Samuelson, Dunleavy & Dowd, 2013; Patterson, Ferguson, Norfolk & Lane, 2005). Selection centres have proved effective when screening applicants for employment, and have been introduced relatively recently as a selection method for medical training (Armstrong, 2006; Evgeniou et al., 2013; Lievens & Coetsier, 2002; Patterson et al., 2005). Preliminary data indicate associations between performance in a selection centre and performance in early general practice training in the UK (Patterson et al., 2005).

There are strong links between assessment centre activities and intended jobs, or roles. Selectors analyse job and performance requirements to identify requisite knowledge, skills and attributes and construct appropriate assessment tasks and scoring criteria. Judgements of candidates are based on observed behaviour, in exercises or simulations that typically include group activities, role-plays, structured interviews and/or tests. An assessment centre—comprised of a range of activity 'stations'—may constitute one stage of a selection process; for example, an assessment centre using situational judgement tests and Multi Mini Interviews might follow review of CVs and academic records and precede review of referee reports. Assessment centres are resource intensive, reducing their feasibility for some uses (Monroe et al., 2013).

Many proponents of assessment centres maintain that they can be used to test psychometric qualities—mental processes—such as "the disposition and ability of individuals to share knowledge" (Armstrong, 2006, p. 183). Burgess et al. (2014) describe using an assessment centre to assess "six domains of practice" (p. 3), including communication and interpersonal skills, clinical reasoning, analytical/problem solving skills; organisational/management skills; sense of vocation/motivation; personal attributes (including the capacity for self-reflection and awareness of the impact of cultural issues on delivery of primary health care) and professional/ethical attributes. High fidelity job simulation situational tests "that place the test taker in a situation … simulating a 'real-life' criterion situation" (Lievens & Coetsier, 2002, p. 246) can offer candidates opportunities to perform and be assessed at the highest cognitive levels appropriate to their stage of training. Ponnamperuma (2010). relates this to the 'shows how' and 'does' levels of Miller's Pyramid of Assessment (Miller, 1990). Referencing Bloom's Taxonomy (Bloom, 1956), such simulations could provide candidates with opportunities to demonstrate high level cognitive performance in analysing, evaluating and creating.

Clinical skills can also be observed and assessed in assessment centres. Gardner, Steffes, Nepomnayshy, Nicholas, Widmann and Fitzgibbons et al., (2017) in their study of simulation in selection to general surgery training in the USA, conducted technical skills exercises—knot tying, suturing, airway management and gowning and gloving—that they found to be surprisingly valuable in assessing candidates' non-technical skills. Koczwara et al. (2012) describe three selection centre exercises for entry to UK General Practice training: (i) a group discussion exercise where groups of four candidates were asked to resolve a work-related issue; (ii) a simulated patient consultation in which each candidate, as the doctor, interacted with a 'patient', and (iii) a written exercise in which candidates prioritised a set of work-related issues, justifying the order chosen (Koczwara et al., 2012). Ensuing selection activities included a competency-based structured interview and a medical interview, comprising technical questions relating to clinical practice. Reviewing these activities, Patterson, Ferguson, Norfolk and Lane (2005) found that supervisor ratings of job performance were associated with assessment centre outcomes, corroborating the validity of the assessment centre activities (Patterson et al., 2005).

Assessment centres typically provide candidates with multiple opportunities to "display their knowledge and skills" (Burgess et al., 2014, p. 8), and comply with Gilliland's (1993) observation of the importance of giving candidates the "opportunity to perform" (p. 701) and thereby contribute to candidates' perceived fairness of the process. Reports have shown that candidates commonly consider assessment centres to be fair and job-relevant (Gilliland, 1993, Patterson et al., 2011), however, in a review of selection to Australian General Practice training, Burgess et al., (2014) identified that procedural justice could be increased by providing candidates with better orientation to selection centre processes. Kolk, Born and der Flier (2003) also found that informing applicants of the dimensions [domains] in which they were being tested improved an assessment centre's construct validity without altering performance outcomes. The susceptibility of assessment centre tests to coaching are not yet reported.

Participation in assessment centre activities may give candidates a better feel for an organisation and its values (Armstrong, 2006; Burgess et al., 2014)—invoking formative learning through the selection process and exemplifying the two-way nature of selection described in the person–environment fit framework.

Although considered by many to provide objective measures of performance, appraisals made in assessment centres are predicated on four major assumptions: firstly, that the identified knowledge, skills and attributes are germane to the job or role; secondly, that the assessment activities elicit the desired skills and attributes; thirdly that candidate performance in the activities predicts their behaviour on the job and, fourthly, that assessors are skilled in making their judgements. The effectiveness of assessment centres' contribution to discerning the 'best' candidates for specialty medical training is unproven and will depend on the extent to which the four assumptions are met—which will vary in local implementation.

Ballot, lottery

To combat the vagueries of selection instruments and candidate attributes, it is sometimes proposed that a 'lottery' may be the fairest way to select into medical training. The best–known example of selection to medical training by lottery occurs in the Netherlands. The Dutch complex, staged system obliges secondary school graduates to participate in a lottery whereby a proportion will gain entry to medical training. Candidates are stratified according to their school grade point average (GPA) with each category allocated a percentage likelihood of admission to training (Cohen-Schotanus, Muijtjens, Reinders, Agsteribbe, van Rossum & van der Vleuten, 2006; Eva, 2014; ten Cate, 2007), , . Although a rationale for instigating the lottery system was to promote fairness, by "[ensuring] equal access to medical education for all students meeting basic entrance criteria" (Cohen-Schotanus et al., 2006, p. 1013), the system has been criticised for limiting diversity—as outcomes are strongly determined by academic achievement—and, counterintuitively, for being 'unfair' to high scoring candidates who have been unsuccessful multiple times.

Since 1999, some Dutch universities have supplemented the lottery with additional, criterion-based processes to exercise more control over selection outcomes, implementing selection tests to admit applicants demonstrating certain attributes, and encouraging diversity by promoting nominated minority groups (Coebergh, 2003; Eva, 2014). Studies of selection outcomes have produced contradictory findings—some found that students whose admittance to training included 'active selection' performed better than those admitted by lottery, others found that they did no better. One study found that students who were admitted solely by lottery were more likely to drop out of training than were those whose selection included assessment of cognitive and non-cognitive abilities (Schripsema, van Trigt, Borleffs & Cohen-Schotanus, 2014; Urlings-Strop, Themmen, Stijnen, & Splinter, (2011). This reinforces Searle and McHarg's (2003) criticism of the Dutch lottery for not "[weeding] out the unsuited as early as possible" (p. 459), thereby contributing to attrition and decreasing the efficiency of training.

The Dutch students' union has criticised universities' active selection methods (assessing candidates' attributes) for not being fair and transparent, admonishing that "every course makes up its own criteria [and] the student does not know what is required" (Coebergh, 2003, p. 138). This appraisal emphasises the importance of selection practices conforming to procedural justice principles in order to maximise their acceptability to candidates and selectors.

At the time of writing, selection to Dutch medical school implements a combination of high school scores, lottery, and active selection. Some universities reverted to lottery-only, citing cost–benefit reasons, while other universities admit some students by lottery alone and some by a combination of lottery and active selection. The Dutch lottery system, while superficially appealing, fair and cheap to run may limit diversity among students through an emphasis on academic achievement and may have long term costs as a consequence of substandard student performance and higher rates of attrition.

Curricula vitae (CV)/Resumes

CVs may include GPA, election to Alpha Omega Alpha (AOA), or the quantity of research projects, publications the candidate has undertaken as well as intended "predictors of motor skill and leadership" (Dirschl et al., 2002, p. 266)—such as hobbies, sporting activities, musical activities, volunteer activities and leadership activities (Alterman, Jones, Heidel, Daley & Goldman, 2011; Dirschl et al., 2002, Makdisi et al., 2011). Prager, Myer and Pensak (2010) criticise the CV for its "factual view of an individual's accomplishments" (p. 327) which does not allow the reader "to know the applicant" (p. 327); however, this view shows little appreciation that selection instruments—each suited to assess particular facets of candidates' knowledge, skills and attributes—combine to provide a comprehensive depiction of candidates. When the CV is implemented to score objective achievements, other selection instruments may be adopted to gain insights into more subjective characteristics.

Valuable aspects of the CV include the clarity, and objectivity of the information presented. One UK application form, uses "includes the curriculum vitae (CV), elements of past achievements and clinical experience, [and] demographic information" (Evgeniou et al., 2013, p. 2 of 7). This UK application also includes "focused questions" (Evgeniou et al., 2013, p. 2 of 7). Love, Ronan-Bentle, Lane and Hegarty (2016) recommend a standardised letter of evaluation (SLOE), bearing some similarity to the RACS' structured CV. Both produce a "clear and concise synopsis" (p. 1480) of relevant candidate experiences, and are standardised, providing information that is directly comparable between candidates. However, it is the 'relevance' of scored activities in the CV that determine the worth of this selection instrument. Schaverien (2016) reflects that "numerous studies" (p. 723) found no correlation between candidates' published research and residents' performance nor any study linking previous academic degrees (such as PhDs) with performance in surgical training. These observations are consistent with the findings of the current study, in which the CV was negatively correlated with clinical performance during training.

Interviews

Interviews have been used extensively to gather information and make judgements about applicants for employment and training positions. Traditionally, interviews are conducted between a candidate and one or more interviewers to ascertain information about each candidate's suitability for a position. Interview formats vary: they may be structured—with all questions standardised for all candidates; semi-structured—with some standardised elements; or unstructured—with few, or no consistent elements (Albanese, Snow, Skochelak, Huggett & Farrell, 2003).

Interviews for selection to specialty medical training are considered important as they allow program representatives to meet candidates and to elicit information about their personal attributes (Collins, 2007; Makdisi et al., 2011). However, traditional interviews have been criticised on many fronts: for lack of predictive validity and reliability, for subjectivity, for context–specificity and for being time–consuming and resource–intensive (Albanese et al., 2003; Campagna-Vaillancourt, Manoukian, Razack & Nguyen, 2014; Eva & Reiter, 2004; Kreiter, Yin, Solow & Brennan, 2004; Makdisi et al., 2011; Prideaux et al., 2011; Schaverien, 2016) . Interviews may be susceptible to coaching, whereby candidates are assisted to prepare exemplary responses to common questions. Such responses are likely to score highly but are unverifiable and may not reflect the candidate's suitability for a training program. Coaching not only confounds fairness of scoring, but also may discriminate against those who cannot access coaching services (Griffin, Carless & Wilson, 2012).

In selection to medical training, the duration of traditional interviews is highly variable, although generally cited as at least 20 minutes. For example: O'Brien, Harvey, Shannon, Lewis and Valencia (2011) report interviews of 25 minutes and 40 minutes; Rosenfeld, Reiter, Trinh and Eva (2008) report on 30- and 40-minute interviews followed by 20 minutes for marking and Albanese et al. (2003) mention a marathon "two- or three-hour interview" (p. 318). Lengthy interviews require great commitment from interviewers, who are usually medical specialists with many competing pressures on their time and availability.

Structured interviews have been shown to be more reliable than semi-structured or unstructured interviews in assessing candidates' personal qualities (Albanese et al., 2003). Unstructured interview content varies from one candidate to another and interviewers may be inconsistently influenced by candidates' qualities or be susceptible to bias or preconceptions (Collins, 2007; Prideaux et al., 2011; Schaverien, 2016). This lack of consistency allows subjective decisions to be made, resulting in potentially unmerited and unjustifiable interview outcomes, and counteracting procedural justice.

Despite variability in interview formats, uncertain validity and reliability, and susceptibility to interviewer bias, selection interviews are valued by some as a means to evaluate candidates' personal attributes (Kelly et al., 2014; Makdisi et al., 2011). It is widely agreed that adding structure enhances the merit of the interview in selection to surgical training (Pau, Jeevaratnam, Chen, Fall, Khoo, & Nadarajah, 2013). Collins (2007) proposes that questions are standardised to assess identified attributes, using clearly defined anchor descriptors and that identical questions are asked of all candidates. In RACS' selection, GS changed from a traditional interview to a structured multi–station interview in 2009.

Multi-Mini-Interviews (MMIs)

Multi-Mini-Interviews (MMIs)—also known as multiple-mini-interviews, or multi-station interviews—apply the principles and protocols of Objective Structured Clinical Examinations (OSCEs) in the selection interview context, affording multiple, independent encounters between candidates and interviewers (Cleland et al., 2012; Eva, Rosenfeld, Reiter & Norman, 2004; Roberts, Walton, Rothnie, Crossley, Lyon, Kumar & Tiller, 2008) . These encounters are usually brief—typically lasting from five to ten minutes, and are structured, as each interview 'station' is focussed on eliciting information about specific attributes or skills. Many variations are possible within the format: the number of stations, the number of interviewers per station, the duration of encounters, the objectives, structure and content of stations. MMIs are a relatively recent addition to selection instruments, being introduced in Canada in 2002 (Eva et al, 2004), but their use is growing (Kelly et al., 2014) and several studies have appraised MMI implementation and predictive validity, particularly for medical student selection (for example: Eva et al., 2004, Eva et al., 2009; Lemay, Lockyer, Collin & Brownell, 2007; O'Brien et al., 2011; Reiter, Eva, Rosenfeld & Norman, 2007; Roberts et al., 2008, 2009). Fewer researchers have considered selection to specialty training; however, recent studies have produced results consistent with those cited above. (Dore et al., 2010; Patterson, Rowett, Hale, Grant, Roberts, Cousans & Martin, 2016; Roberts et al., 2014).

Six to ten interview stations per MMI implementation are common. As the MMI process has matured, some investigations into an optimal number of stations have shown that reliability becomes strong at 10 to12 stations (Cleland et al., 2012; Dore et al., 2010; Pau et al., 2013; Yoshimura, Kitazono, Fujitani, Machi, Saiki, Suzuki & Ponnamperuma, 2015). In their study of selection to medical training, Roberts et al. (2008) deduced that 14 stations would be required to achieve a reliability coefficient of 0.8, although in a subsequent study into selection to general practice training Roberts' team estimated that 10 stations would be required to achieve this reliability (Roberts, Clark, Burgess, Frommer, Grant, & Mossman, 2014).

One or two interviewers are usually engaged per station. Researchers tend to agree that reliability is less affected by the number of interviewers per station than by the number of stations (Dore et al., 2010; Knorr & Hissbach, 2014). This may boost the feasibility of implementing MMIs on occasions where limited numbers of interviewers are available, as a given number of interviewers may reliably adjudicate on more stations. Two caveats to this, however, are to recognise that other benefits may be associated with pairing interviewers and that increasing the number of stations beyond that considered optimal would be counterproductive.

The duration of MMI encounters does not appear to critically affect the reliability of MMIs. Interviewers advised Eva et al. (2004) that eight minutes is ample time to assess candidates' performance. Stations of five to six minutes are reportedly reliable (Dodson, Crotty, Prideaux, Carne, Ward & De Leeuw, 2009; Knorr & Hissbach, 2014). The format of seven– or eight–minute encounters with two minutes between stations for candidates to read information about a scenario that will form the premise for the next station's discussion is common (Dore et al., 2010; Eva et al., 2004; Kelly et al., 2014; Patterson et al., 2016; Pau et al., 2013). Eva and Macala (2014) note that longer stations may be disadvantageous if they provide opportunities for candidates to "sway the conversation to issues that are distinct from the intended focus of the interview" (p. 605), implying that time constraints sharpen the focus of the encounters.

MMIs are generally utilised to appraise candidates' non-technical skills and attributes. The format allows for a high degree of flexibility in station content (Eva et al., 2004; Knorr & Hissbach, 2014) and it is possible to introduce knowledge-based or technical assessment stations, although these might diminish the two-way interactions characteristic of mini-interviews. To test co-operation, Rosenfeld et al. (2008) included a station at which two candidates jointly completed a task. Reports suggest that MMIs benefit from developers clearly identifying the attributes that each station will investigate, mapping these into the overall MMI framework and devising appropriate scenarios, questions and scoring criteria (Kelly et al., 2014; Knorr & Hissbach, 2014; Rosenfeld et al., 2008). Several researchers emphasise that questions must allow for a variety of responses—it is common to devise stations that present candidates with dilemmas, for which there may be no single 'best' answer. Candidates may instead be scored on their ability to think logically and communicate their ideas effectively, or to display professional attributes such as ethical behaviour (Eva et al., 2004, Roberts et al., 2014). Considerable skill is required to design and deliver effective MMIs that test identified attributes. Pau et al. (2013) note that "expertise is … necessary in [both] developing the stations and conducting the interviews" (p. 1030).

'Experience-based' and 'situation-based' questions are frequently used in MMIs. Experience-based questions are 'past–oriented', requiring candidates to recall and describe their experiences when responding to statements or questions (for example: 'what did you do when …'); situation-based questions are 'future–oriented', candidates must imagine how they might respond in hypothetical situations (for example: 'what would you do if …'). Experience-based questions rely on reports of past behaviour to predict future behaviour, while situation-based questions offer candidates with limited experience opportunities to express personal attributes (Ellis, West, Ryan & DeShon, 2002; Eva & Macala, 2014; Roberts et al., 2014).

Some studies have reported the reliability of experience-based, behavioural questions and hypothetical, situational questions as “moderate” (Roberts et al., 2014, p. 10 of 11) or “acceptable” (Yoshimura et al., 2015, p. 8). Eva and Macala (2014) found that behavioural questions were better able to “discriminate between applicants” (p. 609), contrasting with the Culbertson, Weyhrauch and Huffcutt (2017) finding, in their study of employment interviews, that situational questions were better predictors of job performance. However, as interviewers cannot directly observe candidates’ behaviour with respect to the scenarios, both experience-based and situation-based questions allow candidates to present unverifiable information. This increases candidates’ opportunities for ‘impression management’ (Culbertson et al., 2017; Ellis et al., 2002; Eva & Macala, 2014), whereby candidates may say what they anticipate assessors would score favourably, rather than what they actually did or are likely to do. An alternative approach to question content and format, enabling assessors to evaluate directly observed behaviour, is to frame questions so candidates must identify key issues and considerations, evaluate options and discuss implications.

Candidate performance is usually rated against variations of Likert–type rating scales, with anchor points and descriptors to guide assessors. For example, Eva et al. (2004) and Roberts et al. (2014) each reported on 7–point scales and Kelly et al. (2014) reported on a 9–point scale. Assessors benefit from training in the processes, the blueprint, the purpose of the station, the nature of the questions, avoidance of common sources of bias or subjectivity, and in using the rating scales (Roberts et al., 2014; Rosenfeld et al., 2008). Griffin and Wilson (2010) found that interviewer bias was reduced more through skills-based training than by information-based training.

Care must be taken to ensure that the content and context of questions and the scoring criteria do not compromise fairness and equity by introducing biases against atypical candidates, particularly on the basis of gender, age, race, culture or socio-economic background (Campagna-Vaillancourt et al., 2014; Kelly et al., 2014; Razack et al., 2014). This may be challenging—a study of selection to a Canadian medical school implied intrinsic bias in their

MMI; researchers suggested that older candidates and those from a 'higher' socio-economic background may have scored better because they had more experiences to draw from and better communication skills to articulate them (Leduc, Rioux, Gagnon, Bourdy & Dennis, 2016). Such bias may be moderated in selection to specialty training, as all candidates expand their life experience during and following general medical training.

The 'acceptability' of MMIs—involving the format, content relevance and delivery protocols—has consistently been appreciated by candidates and assessors (Campagna-Vaillancourt et al., 2014; Dore et al., 2010; Hofmeister, Lockyer & Crutcher, 2008; Kelly et al., 2014). . Both candidates and interviewers have reported MMIs as being less stressful than traditional interviews—candidates find the format less intimidating than facing a panel and interviewers tend to feel less anxious about making unfavourable judgements as theirs comprises one of multiple assessments (Kumar, Roberts, Rothnie, du Fresne & Walton et al., 2009). Some perceived advantages of MMIs for candidates are that they have opportunities to "recover from poor stations" (Eva et al., 2004, p. 319), to "show who you were" (Kelly et al., 2014, p. 272) and to "present their strengths" (Hofmeister et al., 2008, p. 739). Assessors have suggested that, because scenarios are based on authentic situations, candidates are less likely to predict question content or to rehearse responses and will therefore present more 'genuine' responses (Eva, 2004, p. 316; Kumar et al., 2009, p. 364). Some assessors have speculated, however, that there is the potential for communication skills to predominate over other attributes that may be sought (Kelly et al., 2014).

The considerable flexibility and variability in implementing MMIs may limit the generalisability of MMI studies. Eva and Macala (2014) assert that MMIs are a "process rather than a single instrument" (p. 605) with the results of MMI performance being "entirely dependent on implementation" (p. 605); Knorr and Hissbach (2014) concur, proposing that the "wide range of approaches" (p. 1158) and the "great variability" (p. 1164) in MMI designs greatly affect validity and reliability outcomes.

In the context of this study, the RACS GS interviews comprised semi-structured, multi–station assessments of candidates' attributes as they related to the RACS competencies of scholar and teacher, communication and collaboration, management and leadership, health advocacy and cultural awareness, and professionalism as well as on candidates' contribution to GS. Interview protocols and questions could differ annually and between Au and NZ.

Referee reports (RRs)

Referee reports are regularly used as components of selection for employment and for admission to some forms of training. Several selection instruments share common features with referee reports: letters of recommendation, deans' letters, letters of support, peer–ratings, testimonials and some standardised assessments (Jefferis, 2007; Munro, Bore & Powis, 2012) are all implemented to comment on candidates' knowledge, skills or attributes, as judged by those with direct experience of candidate performance. These instruments have been much criticised, however, as being biased—referees are more likely to act as candidate advocates than as objective reporters (Katsufrakis, Uhler & Jones, (2016); Love et al., 2016). Other critical appraisals maintain that referee reports and letters of recommendation demonstrate poor reliability (Cleland et al., 2012), that they often provide incomplete or inaccurate information (Schaverien, 2016), omit negative comments and contain "widely variable content" (Love et al., 2016, p. 1480) and are difficult to compare against each other (Love et al., 2016; Munro et al., 2012). Many of these criticisms relate to unstructured, subjective referee reports, such as letters of recommendation. To address identified shortcomings, more objective, evidence-based models have been explored (Love et al., 2016; Patterson et al., 2016).

Protocols for implementing referee reports impact upon their effectiveness—these include processes for selecting referees, candidate access to the reports and the number of reports required per candidate. When candidates choose their referees, the likelihood of partiality increases—candidates will nominate those who they consider will promote their interests. Whether the reports are confidential or accessible to candidates may influence referees' comments and scores, potentially impinging on the reports' validity. Requiring more, rather than fewer referee reports has potential to increase reliability, but does not address referee failure to submit reports. Although referee reports are usually used as an ‘additive’ component of selection, it has been mooted that, as most candidates score at the high end of the referee report scoring scale, any low scores are likely to indicate extremely substandard performance. This has led investigators to suggest using referee reports to exclude applicants—by rejecting those who score low in this instrument. However, as Munro et al. (2012) identify, this approach may be best suited to situations in which most candidates are very likely to be selected and the objective becomes one of ‘deselecting’ unsatisfactory applicants rather than selecting the best.

Several aspects of referee performance affect the outcomes of these reports. Ideally, referees have direct knowledge of candidates’ performance, make honest, accurate judgements and return the reports to the selection administrators. In medical specialty training, referees are likely to be supervisors (Jefferis, 2007; Munro et al., 2012). Collins (2007) proposes using those with whom the applicant has worked with in the last two years. Munro et al. (2012) reflect that anonymous referees are able to provide “frank evaluations without fear of embarrassment” (p. 13) but, equally, have no responsibility to provide truthful evaluations. This raises the notions of referee accountability and of selectors, rather than candidates, appointing the referees, to increase the reliability of the reports. It also alludes to the conscientiousness of referees—the amount of time and effort they devote to the task, the sincerity with which they make judgements, whether they submit incomplete reports or do not submit the reports at all. Collins (2007) suggests that the length of the reports may impact on completion rates, implying that fatigue may also be a factor.

Inter-rater reliability will affect referee report outcomes and derives from many components involved in making judgements and scoring candidates. For example, referees may be unable to make accurate judgements due to insufficient direct knowledge of candidates, or referees may score inconsistently, or interpret the rating scale differently. Some referees may not use the full scoring scale, or may judge candidates against individual or atypical standards, or may score inaccurately—such as anecdotally occurs when candidate advocates give unwarrantedly high scores—an action termed by Love et al. (2016) as "grade inflation" (p. 1482). Referees—as judges—make important contributions to the reliability of these reports.

The composition—content and format—of the instrument itself affects its validity. Schaverien (2016) suggests that referee reports predominantly review cognitive abilities, however, these reports are well suited to provide assessments of clinical performance, such as clinical skills, surgical competence and the ability to interact with patients and others (Jefferis, 2007). The formats of the instruments guide referees' responses. As noted above, the number of criteria to be rated may affect completion rates, but may also affect the predictive validity and reliability of the reports. Munro et al. (2012), discussing the rating scales used to obtain judgments, observe that the content of the rating scales must also be appropriate.

Like many other assessments, the reliability of referee reports is improved when they are structured. The content and format of the GS structured SET RR is similar to WBAs, although scoring and completion of RRs is more likely to occur considerably after candidates' clinical performance. Despite this delay between performance and assessment, Oldfield et al. (2013) found that structured referee report scores correlated with performance during subsequent surgical residency.

In the context of this study, the RACS GS RR presents an objective, structured assessment of candidates' workplace performance, as judged by their supervisors, in criteria aligned to the RACS competencies of collaboration, communication, judgement and clinical decision making, management and leadership, medical expertise, professionalism, scholar and teacher, and technical expertise (Oldfield et al., 2013, p. 413).

Situational Judgement Tests (SJTs)

Situational judgement tests (SJTs) are written assessments measuring non-cognitive characteristics, particularly analytical and problem-solving skills, professionalism, ethics and clinical reasoning (Patterson et al., 2016; Roberts et al., 2014). The extent to which responses require specific job–related knowledge varies. SJTs present candidates with hypothetical clinical situations and quandaries and a set of predetermined response options. Candidates must nominate their preferred courses of action, usually by ranking a list of alternatives from most– to least–preferred, or by selecting a 'best' and a 'worst' response, or by rating the effectiveness of the options on a scale (Arthur, Glaze, Jarrett, White, Schurig & Taylor, 2014; Cullen, Sackett & Lievens et al., 2006; Patterson et al., 2016; Roberts et al., 2014). Two common question frameworks are 'knowledge'—what *should* you do? —and 'behavioural'—what *would* you do? (Arthur et al., 2014; Cullen et al., 2006; Nguyen, Biderman & McDaniel 2005). SJTs are substantial assessments; researchers commonly report tests of 50–70 questions, entailing response times of two or more hours (Cullen et al., 2006; Husbands, Rodgerson, Dowell, & Patterson, 2015; Patterson, Ashworth, Kerrin & O'Niell, 2013; Patterson et al., 2016; Roberts et al., 2014) .

SJTs are usually straightforward to administer and score; the written multiple response format readily allows computer delivery and marking, contributing to efficiency of processes and minimising subjectivity in scoring. However, as SJTs rely on the clinical veracity of the scenarios and on panels of subject matter experts to devise relevant responses and to agree on the relevance and ranking of courses of action (Bergman, Drasgow, Donovan, Henning & Juraska, 2006, p. 226) they may be resource- and time-consuming to establish and to monitor. It is possible to devise multiple questions per scenario.

SJTs have been used in selection to medical training and some specialty training in the UK and, more recently, in selection to General Practice (GP) training in Australia (Roberts et al., 2014; Patterson et al., 2016). They have been found to predict performance in medical training. Predictive validity between performance in a SJT and in an end-of-training OSCE was observed in studies conducted by Roberts et al., (2014) and by Patterson et al., (2016). These studies also showed that a SJT had common and independent variance with a MMI, supporting inferences that the SJT and MMI assessed similar, but not identical attributes and that their combined use more strongly predicted training outcomes than did either instrument independently. However, Harris, Walsh & Lammy (2015) caution that there is no data to support the view that high scores in SJTs predict professionalism later "in medicine" (p. 42).

There has been interest in the susceptibility of SJTs to coaching. Information is limited regarding how widespread coaching is in selection to medical or specialty training, however, anecdotally, commercial coaching has been identified as unfairly providing advantages to some candidates over those who cannot or who choose not to use these services. Additionally, coaching, typically a narrowly-focussed, short-term activity, targetted at preparing candidates with strategies and responses for particular tests, may conflict with educational goals such as deep learning of ideologies and conduct (Patterson et al., 2013; Stemig, Sackett & Lievens, 2015) and may encourage candidates to provide learned answers that are intended to 'manage' the impression they make, potentially masking their authentic traits. Cullen et al. (2006) and Stemig et al. (2015) identified that a strategy of providing moderate responses in a rating format SJT would allow candidates to score higher than were they to provide extreme responses. Such strategies on the part of candidates diminish the reliability and predictive validity of the assessments.

To counteract unfair access to commercial coaching, Stemig et al. (2015) recommend making organisationally-endorsed coaching freely available to all candidates, via practice questions with optimal responses, and information sessions. It may be argued that 'coaching', in this instance, could more appropriately be termed 'training' or familiarisation with the assessment format. The purpose of such training is to enhance fairness for all candidates rather than to provide a competitive advantage for some.

Certain question formats can be more susceptible than others to coaching. Coaching appears to have greater effect on responses to simple scenarios than on complex clinical dilemmas, on behavioural-based 'would do' questions rather than knowledge-based 'should do' questions, and on simple question formats (such as 'pick the best response') rather than complex, cognitively loaded formats (such as 'rank all five responses') (Lievens, Buyse, Sackett & Connelly, 2012; Nguyen et al., 2005). Testing several domains in complex, authentic scenarios, with cognitively challenging responses may minimise the potential influence of coaching on SJT outcomes (Cullen et al., 2006; Lievens et al., 2012; Nguyen et al., 2005; Patterson et al., 2013).

Responses that are intended to impress assessors rather than reveal candidates' actual attributes—'faking'—may reduce the reliability of SJTs. Consistent with the findings regarding the effects of coaching, Nguyen et al. (2005) identified that the 'would do' question format was more susceptible to faking than was the 'should do' question format (Harris et al., 2015; Nguyen et al., 2005). A study of selection to five health professions conducted by Lichtwark, Henry, Garvey, Najm and Hay (2017) concluded that SJTs might be used to 'screen out' unsuitable applicants, as they were found to differentiate between low scoring candidates.

Researchers regularly report good acceptability for SJTs, implying good face validity (such as Patterson et al., 2011) however, Sharma (2015) identified in a study of 51 final year UK medical students that although some discerned formative, training aspects of these assessments (exemplified in statements such as "It brought up several things that my medical school training had missed" (p. 235), others were highly critical. Detractors expressed cynicism about the relevance of the test, frustration regarding the relevance and arbitrary nature of response options and mistrust about its validity as a test of professional attributes. Sharma also alludes to potential unfairness of SJTs, noting the difficulty in ranking options accurately "when some may be equally appropriate or inappropriate" (p. 237). Patterson et al. (2011) also noted the SJTs' comparatively low face validity from some candidates' perspectives, despite having "the highest criterion-related validity" (p. 295).

Rockstuhl, Ang, Ng, Lievens and Van Dyne (2015) question whether most SJTs do test 'situational' judgement. They argue that the format is likely to obtain candidates' "response judgements [about] the effectiveness of different response options" (p. 464) rather than "situational judgements" (p. 464) that draw on how the candidates "perceive and interpret situations" (p. 464). The response judgements, that most know as 'situational judgements', are predicated on how candidates make sense of a situation—on their comprehension of the "cues and incomplete information” (p. 465) provided in the scenario—but this initial appraisal of *situations* is rarely tested. Rockstuhl et al. (2015) argue that candidates use their situational judgement to make further judgements about the response options. Two–stage scoring in which candidates first consider and appraise a situation, then review courses of action may tap into this concept.

The efficacy of SJTs is a function of their design features (Arthur et al., 2014). As SJTs are developed uniquely for specific situations, evaluation results may not be generaliseable to other settings and each implementation should be individually appraised (Patterson et al., 2013).

Selection instruments and protocols – specific. Summary

This section has focussed on selection instruments used in medical training. Some seemingly objective instruments, such as AOA membership or a ballot, may be less objective than first supposed. Assessments which present candidates with work–related activities or experiences may have synergies with augmenting person–environment fit, but this is predicated on the authenticity with which they invoke the workplace. As with other selection instruments, the worth of CVs in selection is heavily contingent on the relevance of scored activities to subsequent training. Interviews are valued for providing direct contact between candidates and interviewers. Traditional interview formats are liable to result in subjective, inconsistent judgements. Adding structure to interview formats and questions increases their validity and reliability; multi–mini interviews are a flexible format and have been shown to predict performance in training. Unstructured referee reports or letters of recommendation have been much criticised for highly variable, subjective content. Increasing the structure of referee reports increases their overall reliability, although they are still subject to variability in inter-rater reliability, including a propensity for raters to give high scores. Situational judgement tests attempt to increase the objectivity in assessing non-cognitive characteristics, by using a multiple-choice response format. However, these assessments rely on the relevance of scenarios and on the accuracy and credibility of proposed responses. Situational judgement tests have been found to predict performance in medical training.

Questions regarding the authenticity of candidates' responses and the effects of coaching on candidate performance in many of the selection protocols and instruments are of great concern to selectors. Some reactions include devising items and protocols that are intended to be less susceptible to coaching, or 'flattening' the playing field by making coaching available to all potential candidates.

The selection instruments and protocols presented above is not comprehensive, but has been chosen to represent common instruments, highlight pervasive concerns and discuss the particular instruments implemented by RACS. The ways in which selection instruments are employed in local contexts greatly impacts their validity, reliability and feasibility.

Part 5 Comparable studies

A small sample of studies that have used methodologies similar to that in the current study to analyse relationships between scores in selection and scores in assessments is presented here. Other representative studies are referred to throughout the Literature Review. Typically, such studies assess the extent to which performance in one or more aspects of selection predict subsequent performance. Systematic reviews by Roberts et al (2017) and by Patterson et al (2016) emphasise that reviews of individual selection tools can be limited, inconclusive and contradictory. These researchers highlight gaps in the evidence for the effectiveness of combinations of selection instruments. This study evaluated the performance of individual selection instruments and of the 'Total selection score', representing the performance of combined selection instruments.

Carmichael et al. (2005), in their study of orthopaedic surgical trainees' performance in selection activities and examinations, demonstrated that comparing scores in these instruments revealed the types and the extent of relationships between performance in these activities. Eva et al. (2004) and Eva, Reiter, Trinh, Wasi, Rosenfeld & Norman (2009) similarly used multiple mini-interview scores and subsequent assessment scores to evaluate the capacity of the multiple mini-interview to predict medical students' performance.

In their 2002 study of a dental training program, Sandow, Jones, Peek, Courts and Watson (2002) found a statistically significant correlation between low selection scores and poor performance. Sandow et al.'s use of multivariate correlation to analyse the relationships between many admission criteria (including academic scores, interview scores and a dental admissions test) with dental school performance measures is consistent with the methodology used in the current study.

In a comparable study, Wilkinson, Zhang, Byrne, Luke, Ozolins, Parker and Peterson (2008) assessed how well prior academic performance, admission tests and interviews predicted academic performance in the medical program at the University of Queensland. Wilkinson et al. calculated the correlation between student academic performance and selection criteria for three cohorts (2001–2003 entry years), using Spearman's correlation coefficient. The study used multiple linear regression to compare outcome variables: clinical exams, written exams, and ethics exams, against predictor variables: academic GPA, the Graduate Australian Medical School Admissions Test (GAMSAT) and an interview. The researchers found that the selection criteria used by the University of Queensland predicted about 20 percent of examination performance. Wilkinson et al. (2008) reported that the predictive power of interview performance was higher for academic performance at the end than at the start of the medical program and cautioned that "most variation in academic performance is not explained by selection criteria" (p. 351). The researchers propose that their study was strengthened by examining a range of selection criteria in association with each other, not in isolation.

Summary to the Literature Review

From an examination of prevailing research, I contend that selection to specialty medical training is an inexact pursuit which, despite high-stakes consequences, has yet to be reliably standardised. Selectors aspire to admit individuals who will make proficient trainees and practitioners (Carroll et al., 2009; Cleland et al., 2012; Elfenbein et al., 2015; Gallagher et al., 2008; Martin, 1996; Makdisi et al., 2011; Thordarson, et al., 2007); however, selection decisions are frequently expressions of speculative conjecture about candidates' potential for the future. Multiple facets of selection have been studied to elucidate the nature of their contributions and their impact on selection outcomes. Studies predominantly focus on functional aspects of selection and implications for practice. As implementation of selection instruments and processes appears to be highly context–specific, the findings from these studies frequently differ and may not be directly comparable. Some fundamental principles, such as aspiring to fairness may be so pervasive that they are seldom elucidated; in contrast, there is strong interest, and less agreement, regarding key issues such as the goals of selection and the relative importance of designated knowledge, skills or attributes sought in candidates and how to assess these.

Candidates' perceived merit for selection into surgical training is measured and assessed through their performance in selection instruments, but may also take account of social, cultural, racial or other attributes. There is a tension between the importance of diversity, the need to address historic inequities in the composition of the profession, and the notion of merit-based selection. Compensation for disadvantage or inequity in preparatory experiences may be applied in addition to performance in selection instruments. Judgements regarding the intent and implementation of compensatory measures must be communicated to participants. Review of the effects and outcomes of compensatory measures will feed into the cycle of selection development and implementation activities.

Through the literature, I have identified that non-cognitive skills as well as technical motor skill acquisition and proficiency are important to optimal surgical practice and may develop during training. Some trainees may demonstrate great technical aptitude; others may take longer to learn technical skills, but may demonstrate early proficiency in cognitive or other non-technical skills. Trainees will enter surgical training with varying degrees of aptitude across the range of identified competencies and will develop skills at different rates in response to many stimuli and experiences. The current study did not address aptitude beyond that tested in the RACS selection instruments.

One proposal to maximise reliability of selection is to introduce “more robust work-based assessment tools” to assess performance of candidates prior to selection (Collins, 2007, p. 9). Using results from structured WBAs in selection has the potential to enrich the information that candidates bring to selection by adding ‘authenticity’. It may enhance reliability by reducing the time lag between performance and assessment and may reduce pressure on assessors to complete and return reports to selection administrators. However, such assessment would be reliant on assessors cooperating, regardless of whether they had an affiliation to a specialty training program. To optimise this approach, assessment tasks and instruments would need to be predetermined; then it could draw on a model in which candidates instigate assessments to build longitudinal portfolios of their achievements. Such an approach could improve the validity and reliability of assessments and be cost and time efficient for specialty colleges. Further study would be required to determine the extent to which this approach would assist selectors to differentiate and rank candidates.

In this chapter I have described four conceptual frameworks that contribute to our understanding of selection to surgical training, offered insights into key issues that impact on surgical training and selection, analysed general and specific attributes of selection and assessment instruments and protocols, examined eight prevalent selection instruments, and presented a sample of studies that have used methodologies approximating that in the current study. It is important to note the interplay of multiple influences and determinants that affect protocols and outcomes in selection and assessment. It is also relevant that developments and studies in this field are ongoing, incrementally adding to our identification and understanding of practices to enhance selection outcomes.

The primary purpose of this literature review was to create an information base for investigation of the RACS GS selection instruments. It has presented a set of conceptual frameworks, a broad spectrum of selection issues and attributes and the outcomes of relevant research to support the study. This dissertation will integrate the outcomes of the literature review into an evaluation of selection practices for surgical training in Australia and New

Zealand. Without reliable and valid measures of identified attributes, selection to surgical training will continue to be an imprecise endeavour with uncertain outcomes.

CHAPTER 3: METHODS

The rationale for this study was to examine the efficacy of the relationship between selection variables and assessment variables in surgical training. A goal is to optimise selection methods so those with the greatest likelihood of succeeding in surgical training and in the practice of surgery are most likely to be selected into the Royal Australasian College of Surgeons (RACS) surgical education and training (SET) program. The investigator performed a structured review of records and multiple-year analysis of performance in selection and assessment measures for this quantitative study.

Research questions

Main question

What is the relationship between scores at selection and scores in assessments during the first two years of surgical training?

Subordinate questions

The main research question is supported by the sub-questions:

What are the performance characteristics of each of the selection items?

What are the performance characteristics of each of the assessment items?

To what extent do selection scores predict early assessment scores?

Study design

The study is a correlational design, assessing relationships between selection scores and subsequent scores for in training assessment. The study used quantitative data to analyse relationships between scores in selection and scores in assessments during training for a sample of trainees undertaking surgical education and training in General Surgery (GS) at the Royal Australasian College of Surgeons (RACS). Quantitative data drew on surgical trainees' scores in specified selection and assessment items. Scores at each sample point were compared to determine: 1. The strength of interrelationships between scores in each of the selection items; 2. The extent to which selection scores predict assessment scores; 3. The strength of correlations between scores in selection and assessment items; 4. Similarities and differences between three annual cohorts of trainees; 5. Similarities and differences between Australian and New Zealand trainee cohort selection and assessment performance.

A retrospective, longitudinal design was used, tracking GS trainees in three cohorts—selected in 2008-2010—across the first two years of training. Pearson product-moment correlations and regression analyses were used to determine interrelationships between selection instruments and the predictive validity of the selection instruments. Examples of similar design methodology are described in many studies, including those undertaken by Carmichael et al (2005), Dirschl et al (2002), Eva et al (2009), Patterson, Lievens, Kerrin, Munro and Irish (2013), Poole, Shulruf, Rudland and Wilkinson, (2012), Reiter, Eva, Rosenfeld and Norman (2007) and Sladek, Bond, Frost and Prior (2016).

The study population

Trainees enrolled in RACS SET program in GS constituted the study population. Those within the study population are referred to variously as 'applicants', 'candidates', 'surgical trainees' or 'trainees'. The terms 'participants' and 'subjects' are not used as no recruitment was undertaken to establish the study sample. The group of trainees who were selected in any specific year are referred to as that year's 'cohort'.

The study reviewed selection and assessment archival data for three entire cohorts of Australian (Au) and New Zealand (NZ) GS trainees. Data was reviewed for 347 Au and NZ GS trainees. The three consecutive cohorts that formed the basis of the study applied to RACS in 2008 (n = 100), 2009 (n = 107) and 2010 (n = 140), commencing surgical training in 2009, 2010 and 2011 respectively. The trainees were located throughout Au and NZ—in all major cities and in rural and remote locations. At the time of the study most of those selected in 2008 had progressed into their third year of training (SET3), with the 2009 and 2010 cohorts in SET2 and SET1 respectively.

The instruments

The instruments used to ascertain candidate performance in GS selection to surgical training comprised three selection instruments and total selection scores. The surgical trainee performance assessment items comprised three examination items and three work-based assessment items. These are identified in Figure 6. The assessment scores used in the study comprised each trainee's results in: three selection activities—structured Curriculum Vitae (CV), structured referee reports (RR), structured multi-station interviews (Int)—and a total selection score (Total sel); their first attempt at three examinations—the Generic surgical science examination (GenSSE), the Specialty surgical science examination (SpecSSE) and the Clinical examination (CE)—and ratings in four implementations of each of the three in-training and work-based assessment items—the Direct observation of procedural skills (DOPS), the Mini Clinical Evaluation Exercise (MiniCEX) and the End of term assessment (ETA). See Figure 6. Trainee performance scores in all instruments were converted to percentages.

Selection score data was collected in all selection items and assessment performance data was collected for all available assessments. Candidates' scores were calculated for each selection item, were converted to percentages, and were recorded; an aggregate Total selection (Total sel) percentage score was derived for each candidate. Assessment scores were recorded for trainees' first attempts in each of the three examinations (GenSSE, SpecSSE, CE). Work based assessment reports (DOPS, MiniCEX and ETA) for all trainees for each rotation were reviewed and ratings were converted to numeric scores for each report; scores for sub-components within the ETA reports were also calculated. See Table 1 for DOPS, MiniCEX and ETA scoring conversion parameters. These instruments were chosen to provide the data for this study, as they comprise the key, quantifiable indicators that are used in the GS training program to measure trainee performance.

Table 1 *Conversion of assessment ratings to numeric scores*

Assessment	Maximum score	Rating category	Numeric score
DOPS (form A)	30		
		Unsatisfactory	0
		Borderline	1
		Competent	2
		Excellent	3
		Not observed/Not applicable	Maximum score reduced by 3 points
DOPS (form B)	30		
		Unsatisfactory	0
		Borderline	1
		Competent	2
		Excellent	3
		Significant improvement required	1
		Some improvement required	2
		Competent	3
		Not observed/Not applicable	Maximum score reduced by 3 points
MiniCEX	27		
		Unsatisfactory	0
		Borderline	1
		Competent	2
		Excellent	3
		Not observed/Not applicable	Maximum score reduced by 3 points
ETA	142		
		Competencies	
		Not competent	1
		Borderline	2
		Competent	3
		Excellent	4
		Not observed/Not applicable	Maximum score reduced by 4 points
		Essential criteria	
		Unsatisfactory	1
		Satisfactory	2
		Not observed/Not applicable	Maximum score reduced by 2 points

Independent variables.

The independent variables used in this study comprised scores in the three selection instruments—CV, RR and Int—that were used by RACS to rank candidates for offers to enter surgical training. Summed scores in these three selection instruments formed candidates' total selection scores. These selection instruments suited the research model and provided the entirety of available information about candidates' performance in the selection processes. Each selection instrument utilised different methods to identify and score candidates' attributes. The study could therefore assess the relationship between each independent variable and the independent variables. No candidates had missing scores in selection instruments.

Dependent variables.

In accord with literature arising from previous studies, assessments undertaken during training were used as outcome variables. Assessments used in this study were limited to those undertaken in the first two years of training, comprising DOPS, MiniCEX, ETAs and examinations. These assessments are summarised in Figure 6. The RACS GS assessments comprised both formative, work-based assessments and summative examinations. Information from all but one assessment implemented in the first two years of surgical training was used. The assessment not used in the study was a work-based, formative, mid-term assessment (MTA), that was administered part way through surgical rotations. As the content of the MTA and the ETA was identical, and trainee performance in the two assessments reflected this, it was considered that no benefit would attract to reviewing both assessments. It was therefore decided to only include the ETA as it filled a more critical role and had higher completion rates, possibly due to its dual formative and summative aspects.

Comparable studies

Other studies have used similar methodologies to analyse relationships between scores in selection and scores in assessments in primary medical and specialty training. Many representative studies are referred to in the Literature Review.

Selection instruments

- **Structured Curriculum Vitae (CV)**
- **Structured referee reports (RR)**
- **Structured multi-station interview (Int)**
- **Total selection score (Total sel)**

Assessment instruments

- **Examinations**
  - ***Written Examinations***
    - Generic Surgical Sciences Examination score (GenSSE)
    - Specialty specific Surgical Sciences Examination score (SpecSSE)
  - ***Practical Examination***
    - Clinical Examination score (CE)
- **Work-based assessments**
  - ***Direct Observation of Procedural Skills (DOPS)***
    - Direct Observation of Procedural Skills 1 (DOPS1)
    - Direct Observation of Procedural Skills 2 (DOPS2)
    - Direct Observation of Procedural Skills 3 (DOPS3)
    - Direct Observation of Procedural Skills 4 (DOPS4)
    - Direct Observation of Procedural Skills Average score (Average DOPS)
  - ***Mini Clinical Evaluation Exercise (MiniCEX)***
    - Mini Clinical Evaluation Exercise 1 (MiniCEX1)
    - Mini Clinical Evaluation Exercise 2 (MiniCEX2)
    - Mini Clinical Evaluation Exercise 3 (MiniCEX3)
    - Mini Clinical Evaluation Exercise 4 (MiniCEX4)
    - Mini Clinical Evaluation Exercise Average score (Average MiniCEX)
  - ***End of Term Assessment (ETA)***
    - End of Term Assessment Report 1 (ETA1)
    - End of Term Assessment Report 2 (ETA2)
    - End of Term Assessment Report 3 (ETA3)
    - End of Term Assessment Report 4 (ETA4)
    - Average End of Term Assessment Report (Average ETA)

Figure 6. Selection and assessment instruments in which trainee performance was scored

Project approval and ethics approval

The study was conducted with the approval of RACS. In 2009 the RACS Board of Surgical Education and Training (BSET), which monitored and coordinated activities associated with the nine surgical training programs, approved access to trainee records for this study. In the same year, approval was gained for the study from the Victoria University Ethics Committee and the RACS Ethics Committee, which recommended appointment of a RACS representative as a co-supervisor, ensuring that RACS was apprised of project developments and outcomes.

It was not deemed necessary to gain consent from the trainees to review and analyse data pertaining to selection and assessment scores as: all data were archival, held by the Board in General Surgery (BiGS) and by RACS, under whose auspices the study was conducted. I was an employee of RACS at the time of the study; a condition for approval of the study was to de-identify all data when reporting results.

Ethical issues encountered during this study primarily reflected the confidential nature of the data, which were comprised of candidates' scores in selection items, work-based assessments and examinations. Although individuals' names were recorded when data were collected, care was taken to de-identify all data during analysis and reporting. An additional ethical concern has arisen in reporting on the content of the selection and assessment instruments. It is not possible to reproduce detailed interview or examination content in this study as this content is not in the public domain and these selection and assessment items may be re-used subsequent to this report. Articulating the content of these items may adversely affect their future validity, reliability and fairness.

Psychometric characteristics of the instruments

The validity, reliability and predictive capacity of the selection instruments—also called selection tools—used by RACS GS had not been established prior to the current study. This study provides information to contribute to RACS' understanding of the current and potential uses of these selection instruments.

Selection instruments.

Selection processes and instruments were specified in annual GS Au and GS NZ selection regulations, which were subject to approval by the RACS BSET. See Appendix C for *Selection to surgical education and training in General Surgery regulations*, 2008, 2009, and 2010.

Decisions about appointment into RACS GS SET in both Au and NZ were based entirely on candidates' performance in three selection activities: the structured Curriculum Vitae (CV), structured referee reports (RR) and structured multi-station interviews (Int). Each of the selection instruments constituted a defined proportion of the total selection score (Total sel). Although the CV, RR and Int were used in each annual selection process, iterative differences in the content of these instruments and modifications to protocols occurred annually. Differences between the content and implementation of Au and NZ instruments also transpired.

The three selection instruments cumulatively collected information about candidates' experiences and attributes. The CV recorded applicants' self-reported, authenticated biographic information, clinical experience and academic and personal accomplishments (Oldfield, Beasley, Smith, Anthony & Watt, 2013); these were categorised as medical expertise and technical expertise; scholar and teacher; and management and leadership. The RR scored applicants' workplace performance, as judged by their supervisors, in criteria aligned to the RACS competencies of collaboration; communication; judgement and clinical decision making; management and leadership; medical expertise; professionalism; scholar and teacher; and technical expertise. In 2008, GS Au used a fourth selection instrument, the Hospital Assessment Report (HAR). The HAR content was identical to the RR, but required hospital personnel other than surgeons as assessors. The Int assessed candidates' contribution to GS and personal attributes as they related to RACS competencies: communication and collaboration; health advocacy and cultural awareness; management and leadership; professionalism; and scholar and teacher. See Appendix D for examples of the CV and RR.

Applicants were initially scored against criteria in the CV and RR; candidates' percentage scores in these two instruments were combined and those whose scores surpassed a designated minimum score were invited to interview. Candidates' Int scores were combined with the CV and RR scores to form an aggregate total selection score (Total sel). Candidates were ranked by their Total sel, with those achieving the highest scores being offered positions in the GS SET program. The number selected varied each year, with the intake dependent on the number of training positions—or 'posts'—available. The number of available training posts per year was a factor of the number of extant trainees who vacated training positions—usually this was by progression through the program or completing the program; less frequently by withdrawing or being dismissed from the program.

GS allocated proportional 'weightings' to the three selection instruments within ranges specified by RACS' BSET and consistent with other surgical specialties conducting training under the auspices of RACS. All specialties implemented the same three selection instruments, within the weighting ranges indicated in Table 2.

Table 2 *Selection instruments and weightings*

Selection instrument	Weighting
Curriculum Vitae	15%–25%
Structured referee report	35%–45%
Interview	35%–45%

The information to be gathered by each selection instrument, the allocation of a maximum score per selection item and the protocols for implementing selection instruments were determined independently by each surgical specialty, to suit their requirements. Selection instrument weightings implemented by GS in 2008, 2009 and 2010 are presented in Table 3.

Table 3 *GS selection instruments with proportional weightings per year*

	Structured CV			Structured referee reports			**Structured multi-station interview**			**Total selection score**		
Specified weighting range	15%–25%			35%–45%			35%–45%			100%		
Year	**2008**	**2009**	**2010**	**2008**	**2009**	**2010**	**2008**	**2009**	**2010**	**2008**	**2009**	**2010**
General Surgery Au	20%	20%	20%	30%[a]	40%	40%	40%	40%	40%	100%	100%	100%
General Surgery NZ	20%	20%	20%	40%	40%	40%	40%	40%	40%	100%	100%	100%

Note. Data sourced from Selection to surgical education and training in General Surgery regulations (2008, 2009, and 2010).

a. In 2008 the GS Au RR was weighted at 30%, with an additional selection tool, the Hospital Assessment Report (HAR), weighted at 10%. The HAR was not reviewed in this study; instead, the weighting of the 2008 GS Au RR is considered to be 40%.

Structured Curriculum Vitae (CV).

The RACS GS Curriculum Vitae (CV) CV gathered biographic information pertaining to candidates' clinical experience and academic and personal accomplishments. Collins (2007) states that "the purpose of the [RACS'] CV is to enable applicants to provide a synopsis of their qualifications, meritorious performances, appointments and experience in various areas of medical and surgical practice" (p. 11). Collins (2007) recommends that the marks allocated in the CV "should ensure a balance between clinical experience and academic achievements" (p. 12). The precise information gathered in the CV varied between surgical specialties, as did the value placed on each component, the scoring of CV content and the weighting of this instrument within the selection process. Components in the GS CV, in the years under review, included: surgical experience, skills, qualifications, presentations, publications, prizes and leadership. Scoring allocations for these components varied per year. CV forms were included in the application form in 2009 and 2010. CV information for 2008 is included in Appendix D.

Applicants self-reported their biographic information on online application forms. Documented proof of all accomplishments recorded in the CV was required. This objective information, comprising itemised lists of verified activities, was scored according to predetermined scales, with a fixed maximum score allocated to each category. Two assessors scored the CVs, adhering to a scoring rubric. Each CV score was usually calculated as the mean of the two assessors' scores; if there was a disparity between assessors' scores, the GS chairperson re-scored CV. CV raw scores were converted to percentage scores. Any applicants found to provide unverified or false information were removed from the selection process. Information included by candidates in their applications was only scored if it was pertinent to the defined CV categories. No additional information beyond that specified in the CV was accepted. Annual scoring components and maximum allocated scores in the CV are presented in Table 4.

Structured referee reports (RR).

Structured referee reports (RR) were implemented to assess applicants' workplace performance—their "personal attributes, quality of work and suitability for the SET Program" (*Selection to surgical education and training in General Surgery regulations*, 2008, p. 7; 2009, p. 6; 2010, p. 6), as judged by their supervising consultants, in criteria aligned to the RACS competencies of collaboration; communication; judgement and clinical decision making; medical expertise; professionalism; scholar and teacher; and technical expertise. Referees rated applicants against competency statements. RR content and methods of identifying referees were reviewed annually by the (Australian) Board in General Surgery (BiGS) and the New Zealand Association of General Surgeons (NZAGS), with changes subject to approval by RACS' BSET.

Table 4 *CV components and maximum possible scores in 2008, 2009 and 2010*

Competencies, experiences and accomplishments	Maximum score Au and NZ 2008[a]	Maximum score Au and NZ 2009[b]	Maximum score Au and NZ 2010[c]
Medical expertise and technical expertise			
Surgical and medical experience	8 points	5 points	7 points
Skills courses	2 points	2 points	2 points
Scholar and teacher			
Qualifications (higher degrees, beyond initial medical degree)	3 points	6 points	4 points
Presentations	4 points	4 points	5 points[d]
Publications	4 points	4 points	
Prizes and awards	2 points	2 points	2 points
Scholar and teacher (teaching)	-	-	3 points
Management and leadership			
Leadership	2 points	2 points	2 points
Total score possible	25 points	25 points	25 points

a Data sourced from Selection to surgical education and training in General Surgery regulations (2008).
b Data sourced from Selection to surgical education and training in General Surgery regulations (2009).
c Data sourced from Selection to surgical education and training in General Surgery regulations (2010).
d In 2008 and 2009, presentations and publications were scored separately; in 2010 presentations and publications were combined into a single score.

Referees.

During the years under review in the current study, each applicant nominated two supervising consultants per rotation, with whom they had worked during the previous two years (2008) or four years (2009 and 2010). The specialty contacted three of these consultants per applicant in 2008 (Selection to surgical education and training in General Surgery regulations, 2008), and five consultants per applicant in 2009 and 2010 (Selection to surgical education and training in General Surgery regulations, 2009; 2010), sending each referee a structured referee report form to complete. Referees submitted their reports to BiGS, or to NZAGS, who allocated scores, based on referees' ratings. In 2008, applicants required three valid referee reports in order to proceed in the selection process; five valid reports were required in 2009 and 2010. Referees were not trained in assessing candidates, nor in completion of the reports. A summary of requirements regarding numbers of referee nominations and referee reports per year is presented in Table 5.

Table 5 *Number of referees and referee reports used in 2008, 2009 and 2010*

Minimum requirements per applicant	2008[a]		2009[b]		2010[c]	
	Au	NZ	Au	NZ	Au	NZ
Number of nominated referees[d]	8[e]	8	10[f]	10	10[g]	8-10[g]
Number of referee reports[h]	3	5	5	5	5	5

a Data sourced from Selection to surgical education and training in General Surgery regulations (2008).
b Data sourced from Selection to surgical education and training in General Surgery regulations (2009).
c Data sourced from Selection to surgical education and training in General Surgery regulations (2010).
d Nominated by candidates.
e Two supervisors from each rotation in the previous two years.
f Two supervisors from each rotation in the previous four years.
g Up to three supervisors from each rotation in the previous two years.
h Reports received and scored by GS.

RR content and scoring.

In 2008 and 2009, 20 items were scored in the RR; in 2010 this was reduced to 16 items. Although the competencies were maintained, the competency components differed from those used in 2008 and 2009, and the scoring rubric was changed from a 5-point scale to a 4-point scale (*Selection to surgical education and training in General Surgery regulations*, 2010). In 2008, communication, collaboration and professionalism were designated "essential criteria"; applicants who did not achieve at least a "basic" score in these criteria could be excluded from the selection process (*Selection to surgical education and training in General Surgery regulations*, 2008, items 5.18 and 5.19, p. 8). Maximum scores per RACS competency were defined in GS selection regulations in 2008 and 2009 but were not defined in GS 2010 selection regulations. It is noted that in the 2008 RR, the number of items per competency was not consistent with the scoring specified in the *Selection to surgical education and training in General Surgery regulations*, 2008 (clause 5.13, p. 8). The total possible score for the 2008 RR (160 points), however, was consistent with the total score specified in the *Selection to surgical education and training in General Surgery regulations*, 2008. Each candidate's final RR score was calculated as the average of their referees' scores. These scores were converted to percentage scores. A summary of GS RR item scoring is presented in Table 6.

Hospital Assessment Reports (HAR).

Hospital Assessment Reports (HARs) were implemented to assess applicants' workplace performance as judged by assessors other than supervising consultants; these included unit managers, allied health managers and directors of medical services. As HARs were in all other aspects identical to the RRs, were used in 2008 in Au only, and were discontinued in 2009, they are not analysed in this report. Instead, the weighting of the 2008 GS Au RR is considered to be 40% (the RR was designated at 30% and the HAR at 10% weighting in *Selection to surgical education and training in General Surgery regulations*, 2008). This adjustment allows 2008 GS Au RR scores to be directly compared to 2008 GS NZ RR scores and to GS Au RR scores and GS NZ RR scores achieved in 2009 and 2010.

Table 6 *RR items and maximum possible scores in 2008, 2009 and 2010*

Competencies and attributes	Maximum score Au and NZ 2008	Maximum score Au and NZ 2009	Maximum score Au and NZ 2010
Medical expertise	**8 points**	**8 points**	-
Basic science	8 points	8 points	-
Technical expertise	**24 points**	**24 points**	-
Technical ability	8 points	8 points	6 points
Clinical expertise	-	-	6 points
Judgement – clinical decision making	**24 points**	**24 points**	
Assessment history examinations	8 points	8 points	-
Use of investigations	8 points	8 points	-
Diagnosis	8 points	8 points	-
Judgement	8 points	8 points	-
Patient clinical care	8 points	8 points	-
Judgement under pressure	-	-	6 points
Situation awareness	-	-	6 points
Problem solving	-	-	6 points
Decision making	-	-	6 points
Organisation and planning	-	-	6 points
Communication	**16 points**	**16 points**	
Communication with colleagues and team members	8 points	8 points	6 points
Communication with patients	8 points	8 points	6 points
Collaboration	**16 points**	**16 points**	
Co-operation and interaction	8 points	8 points	-
Working relationships	8 points	8 points	-
Respect for others	8 points	8 points	-
Team involvement	-	-	6 points
Leadership	-	-	6 points
Scholar and teacher	**16 points**	**16 points**	
Learning	8 points	8 points	6 points
Teaching	8 points	8 points	6 points
Professionalism	**56 points**	**56 points**	
Self-motivation	8 points	8 points	-
Response to stress	8 points	8 points	-
Performance insight	8 points	8 points	-

Competencies and attributes	Maximum score Au and NZ 2008	Maximum score Au and NZ 2009	Maximum score Au and NZ 2010
Reliability and punctuality	8 points	8 points	-
Ethical knowledge and behaviour	8 points	8 points	-
Accepting feedback	8 points	8 points	-
Professional integrity	-	-	6 points
Legal, ethical and political awareness	-	-	6 points
Personal attributes	-	-	6 points
Total score possible	**160 points**	**160 points**	96 points

Note. **Bold text** indicates data sourced from *Selection to surgical education and training in General Surgery regulations* (2008, 2009). Non-bold text indicates data sourced from RR forms (2008, 2009 and 2010).

Structured multi-station interview (Int).

The interview was intended to identify factors deemed important to the practice of GS, address key RACS competencies and assess the "suitability of the applicant for training", including their ability to interact with others, to contribute effectively as a team member, to act ethically and responsibly, to demonstrate compassion, to communicate effectively, to assimilate and organise information, to present information concisely, to demonstrate commitment to GS and to promote health maintenance (*Selection to surgical education and training in General Surgery regulations,* 2009, p. 8).

Shortlisting for interview.

Not all applicants were interviewed; the number of candidates who were interviewed differed each year, as did the methods for 'shortlisting' candidates for interview. In 2008, candidates who received at least four "satisfactory" RRs (and two "satisfactory" HARs) were invited to interview (*Selection to surgical education and training in General Surgery regulations*, 2008, clause 3.3.1, p. 4 and clause 7.5, p. 11). In 2009, candidates who received at least 64% for the RR were invited to interview (*Selection to surgical education and training in General Surgery regulations*, 2009, clause 3.3.1, p. 3 and clause 6.3, p. 8). In 2010, candidate's scores in the CV and RR were added; candidates whose combined (raw) score for the CV and RR was 30 or more were invited to interview (*Selection to surgical education and training in General Surgery regulations* (2010, clause 5.17, p. 7 and clause 6.3, p. 8).

Interview process, content and scoring.

As noted for the other selection instruments, selection regulations specified interview processes. Au and NZ selection regulations were combined in 2008 and 2009. In 2010, Au and NZ had separate selection regulations. This helped to clarify processes, as interviews were conducted independently in Au and NZ. In the years under review in this study, GS Au interviewed applicants in June in Victoria, Queensland, New South Wales, South Australia and Western Australia. GS NZ also interviewed applicants in June, with interviews held in Wellington. Interviewers were GS consultants; some training was provided for interviewers. Applicants were required to provide identification when they presented for interview. Applicants progressed through interview 'stations'; all candidates were asked the same initial questions per station, then interviewers could prompt applicants for more information.

GS specified the competencies and attributes to be scored at each interview station, scoring 'communication' at all stations. In most instances, communication scores for all stations were averaged to provide a global communication score; in 2009 and 2010, interviews in NZ included a communication score in each station score (see Table 7 and Table 8). In 2008, six sections: general surgical insight and self-motivation; ethics, audit; patient care skills; team skills, and overall interview and communication skills were assessed in one 30-minute interview in Au, and at four stations in NZ. In 2009 and 2010, RACS competencies: scholar and teacher; communication and collaboration; management and leadership; health advocacy and cultural awareness [cultural awareness was included in the NZ Int but not in the Au Int]; professionalism, and contribution to General Surgery were assessed at five 10-minute interview stations in Au, and at four 10-minute interview stations in NZ. Two interviewers conducted each station at all locations. Each interviewer independently scored candidates, using scoring rubrics and criterion statements as standards. Scoring rubrics variously presented statements delineating “poor”, “fair”, “good”, “very good” and “excellent” responses (*Selection to surgical education and training in General Surgery regulations*, 2008) or “unsatisfactory”, “basic”, “intermediate”, “advanced” or “expert” responses (*Selection to surgical education and training in General Surgery regulations*, 2009 and 2010), for which the scoring range was 1–5. Station scores were finalised by consensus between interviewers. These were then combined by BiGS and NZAGS administrators to achieve total Int ‘raw’ scores, which were converted to percentage scores. A summary of the format of GS interviews is presented in

Table 7 and a summary of interview categories is presented in Table 8.

Some information was ambiguous or missing from regulations, making provision of a maximum total interview score problematic for 2008. In the *Selection to surgical education and training in General Surgery regulations* (2008), clause 7.10 (p. 11) specifies: "The interview will consist of (six) 6 sections"; clause 7.13 specifies the scoring range (1–5) per section. Although not specified in the regulations, it is assumed that the maximum possible score for the interview is 30—representing a maximum score of five points for each of six sections. The 2008 Au interview forms support this assumption. The 2008 NZ interview forms are less clear-cut, providing scoring options for the first five sections only—general surgical insight and self-motivation; ethics, audit; patient care skills; and team skills. To be consistent with the regulations, an additional score for overall interview and communication skills is assumed.

Interview format.

Most stations presented a scenario, posing a situation, problem or dilemma; these were followed by questions to elicit candidates' comments regarding particular aspects of the situation. Most stations presented interviewers with a "criterion statement" delineating critical issues to be addressed and/or defining attributes of "suitable applicants" (interview questions, 2009). Some questions offered interviewers 'model answers' with which candidates' responses could be compared. All questions presented a rating rubric to facilitate scoring. In Au in 2008, five scenarios/questions were delivered at a single interview station. All other interviews were multi-station. In 2009 and 2010, Au interviews were conducted at five stations; in all years, NZ interviews were conducted at four stations.

Interview questions were similar, but not always identical, between Au and NZ. In 2008, all six sections were addressed at each interview, but specific questions could differ between regions (*Selection to surgical education and training in General Surgery regulations*, 2008, clause 7.12, p. 11); for example, in 2008, Au interviewers asked two, and NZ interviewers asked three, sub-questions regarding personal insight.

In 2009 and 2010, each of the five Au interview stations focussed on one of the competency areas. Communication and presentation were assessed at every station; cultural awareness was included in the NZ Int but not in the Au Int. In 2009, the majority of questions were identical for GS Au and NZ GS, although the order in which questions were delivered varied. Most panels comprised two key questions. Two differences between Au and NZ delivery of the 2009 interviews were identified: Au panel two addressed communication and collaboration and Au panel three addressed management and leadership; in the NZ interviews, however, these were combined as part A and part B of question four. Au panel four, addressing health advocacy, was presented in NZ with the addition of a third question, regarding cultural awareness.

Table 7 *Format of GS interviews 2008-2010*

	2008[a]		2009[b]		2010[c]	
Criteria	Au	NZ	Au	NZ	Au	NZ
Number of stations	**1**	**4**	**5**	**4**	**5**	**5**
Number of interviewers per panel/station	**2**	**2**	**2**	**2-3**	**2**	**2**
Duration of interview stations in minutes	**30**	**10**[d]	**10**	**10**	**10**	**10**
Number of questions[e] per panel/station	**6**	**-**[f]	**2**	**2**	**2**	**2-4**
Maximum score per panel/station	**25**	**-**[f]	**5**	**10**	**5**	**10**
Maximum score for all stations	**n/a**	**-**[f]	**25**	**40**	**25**	**50**
Maximum score for communication and presentation	**5**	**5**	**5**	**-**[g]	**5**	**-**[g]
Maximum interview score	**30**[g]	**30**[h]	**30**	**40**	**30**	**50**

a Data sourced from *Selection to surgical education and training in General Surgery regulations* (2008).
b Data sourced from *Selection to surgical education and training in General Surgery regulations* (2009).
c Data sourced from *Selection to surgical education and training in General Surgery regulations* (2010).
d Station duration not specified in regulations; total duration = approximately 45 minutes.
e 'Questions' comprise key, initiating questions and follow-up probing questions.
f Not specified in *Selection to surgical education and training in General Surgery regulations* (2008).
g In 2009 and 2010, NZ scored communication at each station; there was no separate communication score.
h Inferred from *Selection to surgical education and training in General Surgery regulations* (2008)

In 2010, the Int questions, scoring guidelines and scoring criteria differed between Au and NZ. Although the competencies to be assessed were identical for each country, the 2010 Int questions presented different scenarios and options per country. Au interviewers were provided with checklists of six or seven specific criteria to be ticked as candidates mentioned them while NZ interviewers were provided with less specific 'issues to be addressed'. As in 2009, the NZ interview included a question on cultural awareness that was not addressed in the Au interviews. Interview categories and scoring are presented in Table 8.

Table 8 *Interview (Int) categories and maximum possible scores in 2008, 2009 and 2010*

Sections/competencies	Maximum score 2008[a]		Maximum score 2009[b]		Maximum score 2010[c]	
	Au	**NZ**	**Au**	**NZ**	**Au**	**NZ**
General surgical insight and self-motivation	5	5				
Ethics	5	5				
Audit	5	5				
Patient care skills	5	5				
Team skills	5	5				
Overall interview and communication skills	5	5				
Scholar and teacher			5	5	5	10
Communication and collaboration[d]			5	5	5	10
Management and leadership[d]			5	5	5	10
Health advocacy			5	-	5	-
Health advocacy and cultural awareness			-	5	-	10
Professional and contribution to general surgery			5	5	5	10
Communication and presentation			5	5	5	-
Total score possible	30	30	30	40	30	50

a Data sourced from Selection to surgical education and training in General Surgery regulations (2008).
b Data sourced from Selection to surgical education and training in General Surgery regulations (2009).
c Data sourced from Selection to surgical education and training in General Surgery regulations (2010).
d In 2009, NZ interview question (Question 4) combined communication and collaboration, management and leadership. The question was delivered in two parts. Question content sourced from 2009 Selection – General Surgery training programme, Interview questions.

Total selection score (Total sel).

Candidates' scores in each selection instrument were converted to percentages and weightings were applied in the ratio CV 20%: RR 40%: Int 40% to give an aggregate total selection score (Total sel). Applicants were ranked according to their Total sel and offers of positions on the training program were made, based on candidates' national rankings and, in Au, consideration was given to candidates' nominated regional preferences.

Summary of selection instruments and processes, 2008, 2009 and 2010.

Selection processes were defined by RACS-approved selection regulations, which were modified annually. Although there was considerable similarity between Australian and New Zealand selection processes, variations between the composition and implementation of the instruments, particularly of the Int, were evident. Scores in the CV, the RR and the Int were converted to percentages, weighted at approved proportions and combined to achieve a Total sel score. The Total sel score was used exclusively to rank candidates for selection to GS SET. A summary of selection instrument formats and implementation procedures is presented in Figure 7.

Structured CV (20%)

1 per candidate
Online, structured form
Submitted by applicants to surgical specialty
Scoring criteria identified; no additional information scored
Two GS assessors scored CVs independently; board chair re-scored if scores diverged.

Structured referee report (40%)

3–5 per candidate
Online, structured form
Submitted by referees
8–10 referees nominated by applicant (supervisors from past 2–4 years)
Referees selected by specialty from those nominated
RR score represents the mean score of all referees

Structured multi-station interview (40%)

1–5 stations
GS Au and NZ each nominated the number of stations, duration per station, interview questions and the number of interviewers per station.
2008 Au: 1 station
2008 and 2009 NZ: 4 stations
2009 Au and 2010 Au and NZ: 5 stations
Station duration:
2008 Au: 30 minutes
2008 NZ, 2009 Au and NZ, 2010 Au and NZ: 10 minutes
2 interviewers per station
Interviewers scored independently; station scores were consensus by station interviewers
GS Au interviews held in all states; GS NZ interviews held in Wellington

Figure 7. SET GS selection instruments 2008-2010 – summary of formats and implementation procedures.

Assessment instruments: Examinations and work-based assessments.

Decisions about progression through RACS GS SET in both Au and NZ were influenced by trainees' performance in examinations and work-based assessment (WBA) activities. Of interest in this study are the examinations and WBAs undertaken in the first two years of SET: the Generic surgical science examination (GenSSE), the Specialty specific surgical science examination (SpecSSE), the Clinical examination (CE), the Direct observation of procedural skills (DOPS), the Mini clinical evaluation exercise (MiniCEX), and the End of term assessment (ETA). Implementation of the examinations was specified in RACS policy (see Appendix B) and implementation of the work based assessments was specified in annual GS Au and GS NZ training regulations. Incremental differences in the content of these instruments and modifications to implementation processes occurred annually.

The three examination instruments collected information about trainees' knowledge, skills and attributes, particularly as aligned with RACS competencies of communication; medical expertise. The WBAs collected information about technical expertise; communication; professionalism; judgement – clinical decision-making.

Examinations.

The assessments implemented in SET reflected the range of knowledge, skills and attributes required of surgical trainees. All GS SET trainees were obliged to undertake three specified examinations—sometimes called the 'early examinations'—during the first two years of training: a Generic Surgical Sciences Examination (GenSSE), a Specialty-specific Surgical Sciences Examination (SpecSSE) and a Clinical Examination (CE). The written GenSSE and the practical CE were common to all surgical specialties and were considered to be 'generic'. The content of the SpecSSEs was unique to each specialty—GS administered a SpecSSE with questions aligned to knowledge pertaining to general surgery.

Examinations were implemented as summative assessments, although, if necessary, trainees could attempt each of the early examinations up to four times within the first two years of training. If a trainee failed any of these examinations after four attempts, or did not attempt the examinations within the first two years, they were dismissed from the SET program (see RACS policies in Appendix B regarding conduct of the examinations).

The content of each delivery of each of the early examinations was unique—questions were compiled specifically for a single delivery of each examination. Identical examinations were delivered in Au and NZ. The College regulated the number of questions, the weighting of components within the examinations and the format of the questions. Content included a proportion of new questions and some questions that had been used previously, thereby maintaining currency of the examinations while also allowing trainee performance data to be compiled for particular questions and facilitating calibration of examination difficulty levels.

The written GenSSE and SpecSSE tested trainees' knowledge of the surgical sciences of anatomy, pathology and physiology; the SSEs tested trainees' knowledge, understanding and application of these surgical sciences in health and disease as they applied to generic and specialty-specific situations. The CE tested candidates' application of knowledge and performance of clinical skills in a multi-station examination setting. A summary of the early examination formats is presented in Figure 8.

Generic SSE

1 examination

2 days, one 2½-hour written paper each day. Total 5 hours.

Day 1

2008–2011: 100 multiple choice anatomy questions (MCQs)

2012–current: 80 MCQ anatomy questions and 20 image-based anatomy questions

Day 2

60 pathology questions and 60 physiology questions

Specialty-specific SSE

1 examination

1 day x 2½-hour written examination, MCQ format, 120 questions (anatomy, pathology and physiology)

Clinical Examination

1 examination

1 day

2–to 3–hour practical examination, OSCE format

16 stations

Figure 8. SET Surgical Science Examinations and Clinical Examination formats

Generic Surgical Science Examination (GenSSE).

GenSSE – Format and content.

The Generic surgical science examination (GenSSE) was compulsory for RACS trainees in all surgical specialties. The GenSSE, presented as a written, predominantly multiple-choice-format question (MCQ) examination of 220 items, and tested trainees' knowledge of three generic surgical sciences—anatomy, pathology and physiology. Each implementation of the examination comprised a unique set of questions. Most questions were selected from a question 'bank' with some new questions being developed for each implementation of the examination. Before 2008, this examination, previously known as the 'Part 1 Examination' and the 'Basic Sciences Examination', was a three-day examination, comprising 360 MCQs in the surgical sciences—or 'disciplines'—anatomy, pathology and physiology. In 2008 the Basic Sciences

Examination was split into generic and specialty-specific components—the GenSSE and the SpecSSE. In the years 2008–2011, the GenSSE presented 100 questions in anatomy and 60 questions each in the disciplines of pathology and physiology. From 2012 the anatomy component was changed to 80 MCQs and 20 image-based "spot test" questions. Each anatomy spot test question consisted of up to four components to be answered, with each response being no longer than eight words. The MCQ paper and anatomy spot test questions were conducted over two-and-a-half hours on the first day of the two-day GenSSE; the pathology and physiology MCQ paper was conducted over two-and-a-half hours on the second day.

The GenSSE was conducted twice per year (in February or March and June or July) for SET1 and SET2 trainees; being held concurrently at major city centres in Australia (Adelaide, Brisbane, Canberra, Hobart, Melbourne, Perth, Sydney) and New Zealand (Auckland, Christchurch, Dunedin, Wellington).

Once selected into a specialty training program candidates could attempt the GenSSE; trainees therefore had four opportunities to undertake this written examination, prior to and during the first two years of SET. It was each trainee's decision whether (or not) to undertake the examination at any given opportunity. Trainees could attempt this examination as many or as few times as they wished—within the parameters of the four opportunities in SET1 and SET2—until they passed. Candidates were required to exceed a minimum standard in each of the anatomy, pathology and physiology sections of the examination, at the same sitting, before they were deemed to have passed the GenSSE.

GenSSE – Scoring and setting the pass mark (standard-setting).

Trainees gained points for correct answers; points were not deducted for incorrect answers. Each candidate's MCQ answer sheets were scored by optical scanning against a template of correct answers. A unique pass score was calculated for each implementation of the GenSSE.

After each examination anatomy, pathology and physiology discipline committees checked statistics for each question, generated after the papers were scored. This check detected questions that were ambiguous—thus not able to be 'correctly' answered by candidates—and highlighted any questions where the answer key had been incorrectly entered into the template. Question ambiguity was resolved by deleting ambiguous questions from the examination and re-scoring the papers. Incorrect answer keys were resolved by correcting—called 're-keying'—the template and re-scoring the papers. The final examination results were not generated until after these checks and any ensuing re-scoring occurred. The Examination Committee considered and confirmed the examination results subsequent to these processes.

It was possible to 'pass' each discipline but to fail the overall examination. In each of the three disciplines a minimum standard was required, which was set at a lower standard than the aggregate pass standard—two standard errors below. Because the minimum standard in each discipline was lower than the overall standard, it was possible to satisfy the standard for each discipline and still fail overall. To pass the GenSSE, candidates were required to satisfy the minimum standard in each discipline, as well as reaching the aggregate pass standard. Conversely, candidates who did not meet the minimum standard in one or more disciplines, even if they reached the aggregate pass standard through better performance in other disciplines, were not awarded a pass grade. No candidate could carry forward a pass in anatomy, pathology or physiology from one attempt to a subsequent attempt in the GenSSE.

Reliability estimates of the GenSSE have been consistently high, with small standard errors of measurement. A proportion of the questions in each examination was used in previous examinations; longitudinal performance data were therefore available for these questions, serving as 'marker questions', enabling the RACS SSE and CE Committee to determine the relative difficulty of a given examination in relation to previous examinations. The pass mark for each examination was thus adjusted to allow for slight fluctuations in examination difficulty. Records showed that such fluctuations were small, but not negligible, affecting the outcomes of borderline candidates.

GenSSE – Establishing item difficulty.

At every implementation of the GenSSE, analyses were conducted to establish item difficulty. Difficulty and discriminating indices of each question in the GenSSE were determined after the examination by analysing candidates' performance. The 'difficulty' index of a question—perhaps better termed an 'easiness' or facility index—represents the percentage of candidates who answered that question correctly. When applied to the GenSSE, such analysis showed that questions varied in difficulty, with most questions in the 20% to 80% range of difficulty.

The discriminating index of a question (also known as the biserial-r correlation coefficient) is an estimate of the extent to which candidates' scores for that question correlated with their scores on the whole examination. Although theoretically this value could range from -1.00 to +1.00, it is usually positive for most questions in the GenSSE, as could be expected. If the discriminating index of any question was below +0.20 the question was reviewed to determine whether there were flaws in construction or in keying the question.

RACS held periodic workshops for examiners on constructing MCQs and on the use of performance data. Particular attention was paid to recognising and addressing common construction errors—such as inadvertently providing clues or wording questions ambiguously.

GenSSE – Administration.

The College administered the GenSSE, overseeing development of question content and scoring methods as well as coordinating delivery of the examination and securely storing the results. Anatomy, pathology and physiology discipline committees devised, reviewed and compiled questions for the GenSSE. The scoring system, developed by the College to define passing grades, was based on a Rasch scaling model (Griffin, Wu & Zoanetti, 2004). The score required to pass the GenSSE was set individually for each implementation of the examination according to the difficulty of the items within that examination. See the RACS policy *Conduct of the Surgical Science Examination – Generic Component* for information regarding administration of this examination (Appendix B). This examination, introduced in 2008, combined with the SpecSSE to replace the Surgical Sciences Examination (SSE) that had previously replaced the Basic Surgical Sciences Examination (BSS). Trainees who had satisfied the College requirements for the SSE or BSS were not required to undertake the GenSSE.

Specialty Specific Surgical Sciences Examination (SpecSSE).

SpecSSE – Format and content.

Specialty specific surgical science examinations (SpecSSE) were implemented for trainees in all surgical specialties; the content of these examinations differed per specialty. The GS SpecSSE was presented as a written, multiple-choice-format (MCQ) examination of 120 items, testing trainees' knowledge of anatomy, pathology and physiology, as they related to each trainee's surgical specialty. The SpecSSE was introduced in 2008 and was conducted twice per year (in February or March and June or July) for SET1 and SET2 trainees, in tandem with the GenSSE; being held concurrently at major city centres in Australia (Adelaide, Brisbane, Canberra, Hobart, Melbourne, Perth, Sydney) and New Zealand (Auckland, Christchurch, Dunedin, Wellington).

Once selected into a specialty training program candidates could apply to present for the SpecSSE; trainees therefore had four opportunities to undertake this written examination during the first two years of SET; it was each trainee's decision whether (or not) to undertake the examination at any given opportunity. Trainees could attempt this examination as many or as few times as they wished—within the parameters of the four opportunities in SET1 and SET2—until they passed.

GS determined the format and content of the SpecSSE its trainees undertook. In the years of this study, General Surgery and Urology shared a common 'specialty-specific' SSE; trainees in these specialties all undertook the same 120-question specialty-specific MCQ examination under the auspices of RACS. Forty questions in each discipline—anatomy, pathology, physiology—were presented. Each implementation of the examination comprised a unique set of questions. Most questions were selected from a question 'bank' with some new questions being developed for each implementation of the examination.

SpecSSE – Setting the pass mark (standard-setting).

A pass score for each sitting of the SpecSSE was calculated; there was no requirement to achieve minimum scores in any component to pass the examination. The 'bookmark' standard-setting method (Karantonis & Sireci, 2006) was used to set the pass mark following each sitting of the SpecSSE. Panels of four to eight subject experts who were familiar with the expected standards of performance of GS candidates at SET1 and SET2 set the pass mark following each sitting of the SpecSSE. Panel members first clarified standards required of a minimally competent candidate (MCC)—a candidate who met the standard, though barely. They then reviewed a selection of 60 items from the most recent examination to determine the likelihood of a MCC answering the questions correctly. The items were presented in an ordered item booklet (OIB), ranked in order from the easiest item to the hardest item in accordance with how candidates had performed on those questions in previous Basic Science Examinations and GenSSEs. New questions were not included in the difficulty ranking.

Subject experts worked through the samples of examination questions to ascertain the types of knowledge, skills and abilities required to answer each question correctly. Once the questions were reviewed, each panel member placed a 'bookmark' at the question or range of questions where they judged that a MCC would no longer have an 80% chance of answering correctly. As a group, the experts then discussed the reasoning behind their choices of bookmark location. Panel members could then reposition their bookmarks before being provided with impact data—identifying the cut score and pass rate. There was a final round of discussion and bookmark placement before final bookmark placements were recorded. There was no requirement for consensus among panel members regarding the placement of the bookmarks. Each panel member contributed their opinion regarding where the MCC would no longer have an 80% chance of answering correctly. The median bookmark score was derived and was used to calculate the cut score for the exam, which was set at the lower end of a defined 'professional decision zone'. No candidate could carry forward a pass in anatomy, pathology or physiology from one attempt to a subsequent attempt in the SpecSSE.

SpecSSE – Administration.

The College administered the SpecSSE in collaboration with the surgical specialties—RACS oversaw the specialties' development of question content; the College monitored specialties' scoring and 'standard setting' (setting the pass score for each implementation of the examination); RACS also coordinated examination delivery and the secure storage of results. See the RACS policy for administering this examination (Appendix B). This examination, introduced in 2008, combined with the GenSSE to replace the Surgical Sciences Examination (SSE) that, previously, had replaced the Basic Surgical Sciences Examination (BSS). Trainees who had satisfied RACS' requirements for the SSE or BSS were not required to undertake the SpecSSE.

Clinical Examination (CE).

CE – Format and content.

The Clinical Examination (CE) was compulsory for trainees in all surgical specialties. This practical examination was presented as an objective structured clinical exam (OSCE), testing candidates' application of basic science knowledge and performance of clinical skills such as communication, clinical examination and history-taking, diagnosis, image evaluation and interpretation, treatment and management. The CE was comprised of 16 stations, four stations in each of four clinical skill domains: physical examination, communication, history-taking and procedure. Examples of assessment tasks include: patient history-taking and examination; demonstration of technical skills; application of basic science knowledge; data acquisition and analysis; and counselling and communication skills.

Trainees individually progressed through the 16 stations, which each presented an activity, scenario, and/or a simulated patient; trainees were assessed in designated clinical competencies per station. The stations were each attended by one examiner; all trainees progressed through the same stations, being allocated five minutes at each. See Appendix B for *Instructions to Candidates Presenting for the Clinical Examination*, which provides more detailed information about the structure, process and scoring of the CE.

The CE was conducted twice per year (in February or March, and in May, June or July) for SET1 and SET2 trainees, in tandem with the GenSSE and SpecSSE. The examination was conducted concurrently in public (teaching) hospitals, in Au and NZ, conditional on there being at least 10 candidates registered for each site. The CE could be offered in: Adelaide, Auckland, Brisbane, Melbourne, Newcastle, Perth, Sydney and Wellington.

Once selected into a specialty training program candidates could apply to present for the CE. Trainees therefore had four opportunities to undertake this practical examination during the first two years of SET; it was each trainee's decision whether (or not) to undertake the examination at any given opportunity. Trainees could attempt this examination as many or as few times as they wished—within the parameters of the four opportunities in SET1 and SET2—until they passed.

CE – Setting the pass mark (standard-setting).

Candidates were scored at each station using a 25-item checklist and a global rating scale. Pass marks were allocated to each station. A unique pass score was calculated for each implementation of the CE. The overall pass mark for each CE implementation was determined from the station pass marks and the standard error of measurement for that examination.

Prior to 2012, the pass mark for the CE was determined using the Borderline Group method. For each station, the mean score of those candidates rated as 'borderline pass' or 'borderline fail' on the global scale was calculated and became the cut score for that particular station. From 2012, the mark required to pass each station was determined by the Borderline Regression method. This was calculated by regressing candidates' 'station scores' against the global scores and setting the cut score at the station score that corresponded to the point midway between the 'borderline fail' and 'borderline pass' scores. For both the Borderline Group and the Borderline Regression methods, the minimum passing score for the whole examination was the sum of the 16 station pass marks plus one standard error of measurement. To pass the CE, candidates were required to pass at least two stations of each type—physical examination, communication, history-taking and procedure—as well as achieving the minimum passing score for the whole exam. Candidates were awarded a pass or fail grade for the examination.

CE – Administration.

The College administered the CE, overseeing development of question content and scoring methods as well as coordinating delivery of the examination and securely storing the results. The CE Committee devised, reviewed and compiled questions for each implementation of the CE. See the RACS policy for administering this examination (Appendix B).

Work-based assessments.

Two work-based assessments (Direct observation of procedural skills (DOPS) and Mini clinical evaluation exercise (MiniCEX)) provided 'snapshots' of trainees' clinical skills in the workplace. The End of Term Assessment (ETA)—also known as the in-training assessment—reviewed trainees' clinical, operative and professional skills over time. Most clinical supervisors who implemented these assessments received training in conducting the DOPS and MiniCEX work-based assessments. Trainee performance was rated against predetermined, descriptive categories by GS supervisors and trainers with whom the trainee worked during each six-month clinical placement, or 'rotation'. RACS developed core, generic items and scoring methods for the DOPS and MiniCEX which GS could use unchanged or modify to their requirements. GS administered workplace assessments—overseeing the assessment delivery and securely storing results.

The reliability and validity of the DOPS, MiniCEX and ETA had not been established at the time of writing. The current study will provide information to contribute to RACS' understanding of the current and potential uses of these assessment instruments. All scores in the formative work-based assessments were converted to percentages. Any missing scores or

N/A scores within the DOPS, MiniCEX and ETA were identified and percentages were calculated to reflect these.

Direct observation of procedural skills (DOPS).

DOPS assessments were implemented as formative assessments of trainee performance of surgical procedures and were usually conducted in operating theatres. GS trainees were required to undertake two DOPS assessments per rotation during SET1 and SET2, but could choose to undergo more than the minimum number of DOPS assessments. Trainees instigated these assessments, and nominated the procedures to be assessed on. Assessors observed and assessed trainees: gaining patients' informed consent for procedures; preparing for and conducting operations or components of operations; completing post-operative documentation, and reflecting on their performance. Assessors also rated trainees' overall, global performance for each procedure. The reports were discussed by trainees and their trainers or supervisors immediately following the procedure for which the trainee was assessed.

The GS DOPS comprised ten assessment items including communication, preparation, aseptic technique, technical dexterity, awareness of own limitations and self-assessment. Assessors rated trainees using a four-category rating scale. Ratings were 'Unsatisfactory', 'Borderline', 'Competent' and 'Excellent'; an additional category, 'Not observed / not applicable', was also available. A variation of the form offered three scoring options for item 10: 'Significant improvement required', 'Some improvement required', and 'Competent'. Supervisors rated trainee performance by placing a tick against the appropriate assessment category for each scored activity. See Figure 9 for the content and rating scales used in DOPS assessments. See Appendix E for an example DOPS assessment form.

For the current study, all supervisor ratings were converted to numeric scores (Unsatisfactory=0, Borderline=1, Competent=2, Excellent=3), with a total possible score per assessment of 30. Numeric scores for item 10 were allocated to give a maximum of 3 points, regardless of which variation of the form was used (Significant improvement required=1, Some improvement required=2, Competent=3). Item and aggregate scores were calculated for DOPS reports and were converted to percentages to facilitate comparison with other selection and assessment items.

<table>
<tr><th rowspan="2">Item number</th><th rowspan="2">Assessment category</th><th colspan="4">Rating scale</th></tr>
<tr><td>Unsatisfactory</td><td>Borderline</td><td>Competent</td><td>Excellent</td></tr>
<tr><td>1.</td><td>Explains the procedure and complications to the patient and obtains patient's informed consent</td><td></td><td></td><td></td><td></td></tr>
<tr><td>2.</td><td>Prepares for procedure according to an agreed protocol</td><td></td><td></td><td></td><td></td></tr>
<tr><td>3.</td><td>Demonstrates good asepsis and safe use of instruments/sharps</td><td></td><td></td><td></td><td></td></tr>
<tr><td>4.</td><td>Performs technical aspects competently</td><td></td><td></td><td></td><td></td></tr>
<tr><td>5.</td><td>Demonstrates manual dexterity required to carry out procedure</td><td></td><td></td><td></td><td></td></tr>
<tr><td>6.</td><td>Adapts procedure to accommodate patient and/or unexpected events</td><td></td><td></td><td></td><td></td></tr>
<tr><td>7.</td><td>Is aware of own limitations and seeks help when appropriate</td><td></td><td></td><td></td><td></td></tr>
<tr><td>8.</td><td>Completes required documentation (written or dictated)</td><td></td><td></td><td></td><td></td></tr>
<tr><td>9.</td><td>Analyses their own clinical performance for continuous improvement</td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">10.</td><td rowspan="2">Overall ability to perform whole procedure[a]</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Significant improvement required</td><td>Some improvement required</td><td>Competent</td></tr>
</table>

a. Two variations of DOPS forms presented three or four scoring options respectively for item 10.

Figure 9. DOPS assessment items and rating scale

Mini clinical evaluation exercise (MiniCEX).

Mini clinical evaluation exercise (MiniCEX) assessments were implemented as formative assessments to rate trainee performance of competencies essential to the provision of good clinical care and to facilitate feedback. GS trainees were required to undertake two MiniCEX assessments per rotation during SET1 and SET2, but could choose to undergo more than the minimum number of MiniCEX assessments. Trainees instigated these assessments, and could nominate clinical sessions during which they would be assessed, but they were unable to nominate individual patients or conditions; patient encounters were unplanned as they were contingent on clinical outpatient appointments. Assessors observed trainees interact with patients during clinical encounters in the workplace; trainees would typically take a patient history, perform a physical examination and discuss a management plans with patients. Trainees were rated on these activities and on their communication, professionalism and organisation and were also given an overall rating for the encounter. The reports were discussed by trainees and their trainers or supervisors immediately following the clinical examination for which the trainee was assessed.

The GS MiniCEX comprised nine assessment items including history taking, examination, communication, diagnosis and management, professionalism and organisation. Assessors rated trainees using a four-category rating scale. Ratings were ‘Below expectations for level of training’, ‘Borderline’, ‘Competent’ and ‘Excellent’; an additional category, ‘Not observed / not applicable’, was also available. Supervisors rated trainee performance by placing a tick against the appropriate category. The item ‘overall clinical care’ was roughly equivalent to a global rating. See Figure 10 for the content and rating scale used in MiniCEX assessments. See Appendix E for an example MiniCEX assessment form.

For the current study, all supervisor ratings were converted to numeric scores, (Below expectations for level of training=0, Borderline=1, Competent=2, Excellent=3), with a total possible score of 27. Item and aggregate scores were calculated for MiniCEX reports and were converted to percentages to facilitate comparison with other selection and assessment items.

Item number	Assessment category	Rating scale			
		Below expectations	Borderline	Competent	Excellent
1.	History taking				
2.	Physical examination				
3.	Communicates to patients (and their family) about procedures, potentialities, and risks to encourage their participation in informed decision making				
4.	Adjusts the way they communicate with patients for cultural and linguistic differences and emotional status				
5.	Recognises what constitutes ‘bad news’ for patients (and their family) and communicates accordingly				
6.	Recognises the symptoms of, accurately diagnose, and manage common problems				
7.	Professionalism				
8.	Organisation/efficiency				
9.	Overall clinical care				

Figure 10. MiniCEX assessment items and rating scale

End of term assessment (ETA).

End of term assessments (ETAs) were implemented as assessments of trainee clinical performance throughout six-month rotations. GS trainees were required to undertake one ETA per rotation throughout SET training (from SET1 to SET5 or SET6); trainees therefore were usually assessed twice per year using this method. ETAs had both formative and summative assessment roles. Formative aspects included identifying trainees’ strengths and weaknesses and providing a basis for discussions between supervisors and trainees to guide future development. The major summative roles of ETAs were as ‘sign off’ for satisfying, or not satisfying the requirements of each rotation, authorising trainees to progress to the next stage of training or constraining them to repeat a stage, to undertake remedial activities or, in extreme cases, being dismissed from the training program. Drawing on information from specialty supervisors and trainers, with particular emphasis on the most recent 3-months, these reports were discussed by trainees and their supervisors at one-to-one 'End of Term' meetings.

GS ETAs comprised ten categories (one for each of the nine RACS competencies, plus "essential criteria") incorporating 40 assessment items. GS supervisors rated trainee performance in specified competencies using a four-category rating scale (N: Not Competent, B: Borderline, C: Competent, and E: Excellent), and rated trainees in essential criteria using a two-category scale (U: Unsatisfactory and S: Satisfactory). For the current study, all supervisor ratings were converted to numeric scores, (N = 1, B = 2, C =3, E = 4, U = 1, and S = 2), with a total possible score of 142 for the ETA. See Table 1 and Table 9. Scores for each ETA category and aggregate scores were calculated and converted to percentages to facilitate comparison with other selection and assessment items. See Appendix E for an example of an ETA assessment form.

Table 9 *ETA items and maximum possible scores*

Category	**Maximum score**
Competencies	
Medical expertise	4
Technical expertise	28
Judgement	32
Communication	8
Management and leadership	4
Collaboration	12
Health advocacy	8
Scholarship and teaching	8
Professionalism	20
Essential criteria: Communication, co-operation, self-motivation, work ethic, ability to manage stress, honesty, empathy, teamwork, insight/self-awareness	18
ETA total	142

Summary of assessment instruments and processes.

RACS and GS collaborated closely to assess trainees. SET assessment processes were defined by RACS-approved regulations, which could be reviewed and modified as required. Australian and New Zealand GS assessments were identical. Examinations were closely monitored for reliability and face validity; their primary role was summative. There is no evidence of work-based assessments being reviewed for reliability—although, being implemented in trainees' workplaces as observed assessments of actual clinical performance, face validity is likely to be strong. The DOPS and MiniCEX were targetted, formative assessments; the ETA was more comprehensive, with both formative and summative aspects. All assessments except the ETA assessed trainee performance at a single implementation; ETAs were used to assess trainees' performance over the course of 6-month rotations. For this study, assessor ratings in the DOPS, the MiniCEX and the ETA were converted to percentage scores to facilitate comparison with other assessments. A summary of assessment instruments and formats is presented in Figure 6. The end of term assessments (ETA) assessed six competencies in common with the RRs. See Table 10.

Table 10 *RACS competencies assessed in RRs and ETAs*

Competencies and attributes	Max RR score 2008 2009	Max RR score 2010	Max ETA score 2008 2009 2010
Medical expertise	8	-	4
Technical expertise	24	12	28
Judgement – clinical decision making	24	30	32
Communication	16	12	8
Management and leadership	-	-	4
Collaboration	16	12	12
Health advocacy	-	-	8
Scholar and teacher	16	12	8
Professionalism	56	18	20
Essential criteria	-	-	18
Total score possible	160	96	142

Procedures

Pilot study.

In 2010 and 2011 a pilot study was undertaken of 22 Paediatric Surgery (PS) trainees to test the proposed methodology with a small sample of trainee records. The Chair of the RACS' Board of Paediatric Surgery and the Paediatric Surgery Executive Officer facilitated access to the records. Paediatric trainees' scores for three selection cohorts were extracted and compared to examination and work-based assessment scores using SPSS 19.0. Analysis of the data was conducted using correlation analysis and multiple linear regression analysis. This analysis confirmed the methods were appropriate to extend to a full-scale study.

Full study.

Following the pilot study, a full-scale study was undertaken. This study examined relationships between scores at selection and scores in early examinations and work-based assessments for all RACS GS trainees who were selected to training in 2008, 2009 and 2010 and commenced training in 2009, 2010 and 2011. The study was undertaken in two parts. In the first part, intra-assessment associations, within each of the assessment categories were reviewed. Pearson product-moment correlation analyses (Walker, 2008) were conducted to test associations for trainee performance in selection, in exams, in DOPS, in MiniCEX, and in ETAs. In the second part of the study, performance in selection items was compared to performance in assessment items. Pearson correlations were conducted to determine the degree of association between performance in the selection items and performance in each of the subsequent assessment items. Multiple linear regression was employed to help determine which, if any, of the three selection items and/or the total selection score, could be used to predict the outcomes of assessments for Australian and New Zealand GS surgical trainees.

Reporting.

To maintain the confidentiality of the selection and assessment scores, all data was de-identified before being reported in the study. Data analysis used combinations of grouped scores to identify trends; it is therefore not possible to identify any individual's scores. Trainee names and the unique identifying number (iMIS ID) allocated to all applicants by RACS, which were used in initial data collection and analysis, are not used in reporting the results.

Accessing and compiling archival data

Some of the variables investigated in this study were available as electronic records; other variables, such as measures of work-based performance, were extracted from hard copy, administrative files. Data pertaining to performance in selection measures were gathered and analysed annually by RACS for internal review purposes; these were stored as electronic records. Data pertaining to examination measures were similarly retained and analysed biennially. At the time of the study, data pertaining to performance in work-based assessments were retained in paper form in trainee files. Approval to access data for this study was obtained from RACS, through its Board of Surgical Education and Training (BSET).

The Board in General Surgery (BiGS) and the New Zealand Association of General Surgeons (NZAGS) facilitated access to GS trainee files. Original selection data were accessed in electronic form from spreadsheets provided by BiGS. Original work-based assessment data were accessed from trainee files. Alpha grades for work-based assessments were expressed as numeric scores and were transcribed to electronic spreadsheets using Microsoft® Excel®. Original examination data were accessed from the RACS electronic database, iMIS 15. All scores provided in alpha grades or number form, were converted to percentage scores.

Australian (Au), New Zealand (NZ) and combined Australian and New Zealand (ANZ) GS data are presented for all tests. The combined ANZ results primarily reflect the Australian data, due to the greater proportion of Australian trainees; similarly, the Total selection scores (Total sel) represent the proportional weightings of the selection items (CV 20%, RR 40% and Int 40%). The study uses performance data for three cohorts of trainees, selected in July 2008, 2009 and 2010, commencing SET in December the same year (in NZ) and in January the following year (in Au).

The work-based assessment data was collected in early 2012 and does not include 2012 DOPS, MiniCEX or ETAs, although 2012 examination performance is included. All raw scores have been converted to percentages.

Type of data.

The data collected comprised scores for performance in four selection items and in six assessments undertaken during training. Archival data were sourced from trainee records held by BiGS, NZAGS and RACS. Trainee name, ID number, year of application, country- and state-of-origin and selection scores were collected for all trainees in the study. Performance data for selection items—CV, RR, Int and Total sel—were collected for all candidates. Performance data for assessments—DOPS, MiniCEX, ETAs and examinations—were not comprehensive. At the time of data collection many trainees selected in 2009 and 2010—commencing SET in 2010 and 2011—had not yet undertaken assessments that were implemented in SET2 and SET3. Some trainees did not undertake all the assessments due to exemptions, non-compliance or in anticipation of later implementations of these assessments. Where available, data were collected for every trainee for every work-based assessment and for the trainee's first attempt at each examination. Subsequent examination attempts (trainees who failed an examination were permitted up to three further attempts) were not included in this study. Table 11 presents a summary of the data collected and the selection items and assessment items in which trainee performance was scored.

Table 11 *Data collection*

Demographic information	Data collected
Candidate name	
ID number	Unique number, allocated by RACS
Country	Australia; New Zealand
State (if Australian)	New South Wales; Queensland; South Australia; Tasmania; Victoria; Western Australia
Application/selection information	
Year of application	2008; 2009; 2010
CV	Score
Referee report	Score; sub scores
Interview	Score; sub scores
Total selection	Score
Assessment information	
Examinations	
GenSSE (First attempt)	Score; date; result (Pass or fail)
SpecSSE (First attempt)	Score; date; result (Pass or fail)
CE (First attempt)	Score; date; result (Pass or fail)
Work-based assessments	
DOPS (All undertaken)	Score
MiniCEX (All undertaken)	Score
ETA (All undertaken)	Score; sub scores

All 'raw' examination scores were converted to percentages. All the non-numeric ratings from work-based assessments—DOPS, MiniCEX and ETAs—were converted to numeric scores and percentages. The data, including trainees' scores and sub-scores, where available, for each of the selection items and each of the assessment items, were collated into spreadsheets using Microsoft® Excel®.

Selection scores.

Applicants' selection scores were retained in computer-based records (as Excel® spreadsheets) by the BiGS. Trainee selection data for this study was obtained in the greatest detail available.

Assessment scores – Examination scores.

Examination scores were retained by RACS (Examinations Department), as components of trainee activity records. This data was accessed from the RACS electronic database (iMIS 15). Examination performance data—raw scores, percentages, pass or fail results—are retained for all attempts at the examinations. Data for this study were limited to trainees' first attempt at each of the GenSSE, SpecSSE and CE.

Assessment scores – Work-based and in-training assessments.

Completed work-based and in-training assessment reports—DOPS, MiniCEX and ETA—were retained by BiGS, with regional records maintained locally in regional offices in Adelaide, Brisbane, Melbourne, Sydney and Wellington. At the time of data collection BiGS only retained paper copies of trainees' workplace assessment reports. For ready access to the workplace assessment data in the current study the paper reports were all scanned and stored electronically as portable document format (pdf) files.

All SET trainees were obliged to participate in three types of work-based and in-training assessments during the first two years of training; these were the Direct Observation of Procedural Skills (DOPS), the Mini Clinical Evaluation Exercise (MiniCEX) and the End of Term Assessment Reports (ETAs). The content of generic DOPS and MiniCEX were developed by RACS; surgical specialties could modify these to suit their requirements; the content of the ETAs was unique to GS. BiGS implemented these formative, practical assessments in the trainees' workplaces. Most training and work-based assessment occurred in public hospitals where trainees were employed.

Data collection and collation.

Data were obtained from RACS' records and from GS' records. I had full access to all the records required. Original data were sourced and stored in electronic files, categorised by specialty, selection year, country/state, trainee and their results for selection activities and each of the assessment activities. Paper reports were electronically scanned, converted into pdf documents and stored in electronic files.

Selection scores were obtained from BiGS and NZAGS in the form of electronic Excel® spreadsheets, per selection year. These were stored in electronic files by year and country. Name, ID number, year of application, country- and state-of-origin and selection scores, obtained from the selection spreadsheets were the basis for new spreadsheets into which all data were compiled using Microsoft® Excel®.

Work-based assessments (DOPS, MiniCEX and ETAs) were sourced as original, hard-copy (paper) reports, from trainee files in GS regional offices. I visited GS offices in Au, scanned these reports and stored them electronically as pdf files; administrative staff in NZ scanned reports and sent them to me electronically as pdf files. Reports were individually reviewed: grades for every item within each report were converted to numeric scores and entered into Excel® spreadsheets (categorised by specialty and selection year). Item scores were added, resulting in report scores that were converted to percentages. Each trainee thus had percentage results for every DOPS, MiniCEX and ETA.

Examination results were accessed as electronic pdf reports, generated from the RACS database (iMIS 15). The reports were classified by the year GS trainees commenced SET and listed the examinations that they attempted with their score and outcome (pass or fail) at each attempt.

Compilation.

Descriptive statistics were compiled for all trainees in the study, including the number of individuals, the country selected into, the selection year (cohort) and scores in each of the selection and assessment items. Mean scores and standard deviations were calculated for all selection and assessment items. These data were compiled into Excel® spreadsheets. Where available, sub-scores in each of the items were included. As later data became available from trainees undertaking assessments—such as examination scores and new DOPS, MiniCEX and ETA results—they were included and underwent the same collation and compilation process. Mean scores and standard deviations were re-calculated.

Data analysis

Data were analysed using Statistical Package for the Social Sciences (IBM SPSS) 19.0. Analysis was conducted by the candidate. Data were considered for selection and assessment items for RACS GS trainees for each yearly cohort (2008–2010) separately, for the three cohorts combined and for Au and NZ trainees separately and for both countries combined (ANZ).

Firstly, Pearson product-moment correlations were calculated to examine the extent of relationships within each set of variables. This was undertaken to determine intra-relationships of the selection items—that is, relationships between performance in the CV, RR and Int; of the work-based assessment items—relationships between the performance in DOPS, MiniCEX and ETA; and of the examination items—relationships between performance in the GenSSE, the SpecSSE and CE. All significant relationships ($p < .05$ and $p < .01$) were identified. Many alternative guides have been developed to define the strength of correlations (Eva et al., 2009; Patterson et al., 2013; Poole et al, 2012). In this analysis, the strength of the associations (values of correlation coefficients, expressed as 'r') were considered from 'very weak, negligible' to 'very strong' (Shortell, 2017; "Pearson Product-Moment Correlation", 2017) as indicated in Table 12. The categorisation of the strength of findings was arbitrarily set, as an adaptation of that presented in Mukaka (2012).

Table 12 *Guide to interpreting correlation coefficients (r)*

	Coefficient, (r)	
Strength of association	Positive	Negative/Inverse
Very weak, negligible	**.01 to .15**	**-.01 to -.15**
Weak	**.16 to .29**	**-.16 to -.29**
Moderate	**.30 to .49**	**-.30 to -.49**
Strong	**.50 to .69**	**-.50 to -.69**
Very strong	**.70 to 1.0**	**-.70 to 1.0**

Secondly, Pearson product-moment correlations were conducted to determine degrees of association between selection items and performance in each of the subsequent assessment items. Correlations were calculated to identify associations between scores in selection and scores in assessments for which there was data for more than five trainees. Scores in each selection instrument and the Total sel score were tested against scores in each assessment item and against average scores in assessment categories—examinations, DOPS, MiniCEX and ETAs—by yearly cohort, for the three cohorts combined, by country and for both countries combined. All significant relationships ($p < .05$ and $p < .01$) were identified. As in the intra-assessment analyses, the strength of the associations (values of Pearson correlation coefficients, expressed as 'r') were considered from 'very weak, negligible' to 'very strong' as indicated in Table 12.

Thirdly, multiple regression analysis was conducted to determine the extent to which trainees' scores in the selection items (independent variables) predicted scores in each of the assessments during training (dependent variables). Regressions were conducted for the three selection items—CV, RR and Int—against each assessment—Exams, DOPS, MiniCEX and

ETAs—using a direct method; separate regressions were run for the Total sel against each assessment.

The model fit and strength was assessed using the analysis of variance (ANOVA) step within the regression analysis. The relative strength of the associations between dependent variables and the independent variables were assessed using regression coefficients. Regression values (ANOVA) were considered statistically significant at $p \leq .05$. Results are reported where a) the 'Selection items' model is significant (and 'Total sel' $p > .05$); b) both the 'Selection items' and the 'Total sel' models are significant; c) 'Total sel' is significant and one or more selection items is significant and the 'Selection items' model $p > .05$); and d) the 'Total sel' only is significant. A summary of the statistical analyses to determine the predictive relationship between performance in selection items and performance in assessments is presented in Figure 11.

Descriptive statistics for each of the selection and assessment items, including number of individuals, country selected into, selection year, mean scores and standard deviations in all selection and assessment items.

Pearson product-moment correlations were conducted to examine the interrelationships within sets of variables (i.e. for Selection, to determine the strength of the interrelationships between scores in the CV, the RR and the Int; for Examinations, to determine the interrelationships between the GenSSE, the SpecSSE and the CE; for the In-training Assessments, to determine the interrelationships between the ETA, DOPS and MiniCEX).

Pearson product-moment correlations were conducted to examine the relationships between sets of variables (i.e. between selection and assessments, to determine the strength of the relationships between scores in the CV, the RR and the Int and the examinations and the work-based assessments).

Multiple regression analyses were conducted using Direct (simple) regression to determine the extent to which scores in selection items predicted scores in the examinations and the work-based assessments.

Figure 11. Statistical analyses conducted of GS trainees selected in 2008, 2009 and 2010

Summary to the Methods

A primary goal of the data analyses was to examine the predictive validity of scores achieved in selection to surgical training. This involved the following procedures:

1. Computation of descriptive statistics for measures of selection and assessment performance. Selection measures comprised each of the three selection instruments and the total

selection scores; assessment measures comprised each of the three examinations and each of the three work-based assessments. Descriptive statistics were means and standard deviations.

2. The use of Pearson product-moment correlation analysis to examine the relationships between scores in the selection measures and between scores in assessment measures. Correlations were calculated between selection items and total selection scores and (a) each of the three examinations, (b) each of the three work-based assessments. These included trainees' first attempt at the examinations and scores in every work-based assessment undertaken in the first two years of surgical training.

3. Multiple linear regression was employed to determine which of the selection items predicted performance in assessments. Regressions were calculated between selection items and total selection scores and (a) each of the three examinations, (b) each of the three work-based assessments.

The methods used in this study were consistent with similar studies conducted elsewhere. The methods revealed relationships between selection and assessment items, as reported in the Results chapter. The method could be extended to include results from subsequent examinations and results from trainees in other surgical specialties.

CHAPTER 4: RESULTS

This study examined the relationship between scores at selection and scores in early examinations and work-based assessments. The research questions are:

Main question

What is the relationship between scores at selection and scores in assessments during the first two years of surgical training?

Subordinate questions

The main research question is supported by the sub-questions:

What are the performance characteristics of each of the selection items?

What are the performance characteristics of each of the assessment items?

To what extent do selection scores predict early assessment scores?

The results are reported in two parts. Results are presented in Part one for intra-assessment associations. Results are presented in Part two for performance in selection items compared to performance in assessment items. Australian (Au), New Zealand (NZ) and combined Australian/New Zealand (ANZ) General Surgery (GS) data were analysed for all tests.

Data analyses: General Surgery trainee performance in selection and assessments

Part one of the study reviewed trainee performance within each of the selection and assessment categories. Statistically significant associations ($p < .05$ and $p < .01$) are highlighted in correlation tables. Part two of the study compared trainee performance in selection with performance in subsequent assessments to determine which, if any, of the three selection items and/or the total selection score could be used to predict the outcomes of assessments for Au and NZ surgical trainees.

Part one: Trainee performance within assessment categories – Descriptive statistics and Pearson's correlations

Performance characteristics of selection and assessment items

The study considers candidates who were selected into General Surgery (GS) training, 2008, 2009 and 2010; trainee numbers in these cohorts varied annually. This study used data from trainees who commenced GS training and undertook one or more assessments. Table 13 presents the number of GS trainees in this study.

Table 13 *Number of GS SET trainees per year and country*

Selection year	**Australian (Au) GS trainees**	**New Zealand (NZ) GS trainees**	**Australian and New Zealand (ANZ) GS trainees**
	n	**n**	**n**
2008	81	19	100
2009	89	18	107
2010	121	19	140
Total	291	56	347

Selection items

As identified in Table 37, mean scores for CVs are consistently low (≤54%), the mean yearly CV SD is 12.85, which is greater than either the RR (mean yearly SD = 6.76) or the Int (mean yearly SD = 6.91); mean scores for RRs are typically high (>76); mean scores for Interviews (Int) are also high (>80%) with the exception of NZ in 2010 (73.16%); mean Total Selection (Total sel) scores are high (>73). Total sel SD implies a narrow range of scores, (<4.45) except Au 2008 (SD 6.38). Per cohort, mean Au CV scores are lower than NZ CV scores, mean Au and NZ RR scores are aligned for two cohorts with mean Au RR higher than NZ RR in 2010; mean Au Int scores are all higher than NZ Int scores, and mean Au Total sel scores are substantially aligned with NZ Total sel scores per year, with Au Total sel slightly higher than NZ Total sel, particularly in 2010 scores. The mean score for the RR in Au and NZ increased in 2009 and 2010, as did the mean Total sel score. The SD for the RR decreased in 2009 and 2010. No data were available from RACS on the reliability of the selection instruments.

Examinations

Table 38 shows that the mean scores for GenSSE (range = 72.11–75.86) and the mean scores for the SpecSSE (range = 71.76–76.06) are similar for all cohorts in both countries. The Au and NZ mean scores for the practical CE are also aligned per cohort, and are lower than mean scores for the two written examinations; the 2008 cohort's mean CE scores are lower than the subsequent cohorts' scores. Fewer Au trainees from the 2008 cohort attempted the examination than from the 2009 or 2010 Au cohorts; NZ trainee numbers were stable for the GenSSE and CE across the three cohorts but increased for the SpecSSE (n = 10–18). For all examinations and all cohorts SD range is 4.06 (2009 NZ SpecSSE) – 7.77 (2010 NZ SpecSSE) with most SD within the range 5.00–6.00.

Direct Observation of Procedural Skills (DOPS) assessments

Table 39 shows the mean scores for DOPS. Excluding $n < 5$, DOPS mean scores range from 78.59 (2008 Au DOPS2, $n = 39$) to 85.78 (2008 Au DOPS3, $n = 20$) and DOPS SD ranges from 9.17 (2010 NZ DOPS1, $n = 17$) to 14.18 (2008 Au DOPS3, $n = 20$). Au mean scores in DOPS4 are higher than for DOPS1 in all three cohorts. NZ mean DOPS scores increased from DOPS1 to DOPS2 in all years; however, limited data preclude extending the NZ analysis to DOPS3 or DOPS4. Au and NZ scores are similar for DOPS1 and DOPS2 for the 2009 and 2010 cohorts, however the 2008 Au cohort mean scores for DOPS1 and DOPS2 are lower than those for the NZ cohort.

Mini Clinical Evaluation Exercise (MiniCEX) assessments

The mean scores for MiniCEX, (excluding $n < 5$), range from 79.72 (2010 Au MiniCEX4, $n = 9$) to 91.64 (2010 NZ MiniCEX2, $n = 8$). MiniCEX SD ranges from 8.52 (2010 NZ MiniCEX1, $n = 17$) to 20.02 (2008 Au MiniCEX4, $n = 9$). The 2009 and 2010 Au cohorts showed an increase in mean scores from MiniCEX1 to MiniCEX4; NZ mean MiniCEX scores for all cohorts increased from MiniCEX1 to MiniCEX2 but NZ data are insufficient to be reliable for MiniCEX3 and MiniCEX4. The 2008 Au and NZ mean scores for MiniCEX1 and MiniCEX2 are aligned; the 2009 Au mean scores are lower than the NZ mean scores for MiniCEX1 and MiniCEX2; the 2010 Au mean score is higher than the NZ MiniCEX1 mean score, and lower than the NZ MiniCEX2 mean score. See Table 40.

End of Term Assessments (ETAs)

Mean scores for ETA assessments, excluding $n < 5$, range from 79.30 (2010 NZ ETA3, $n = 7$) to 87.73 (2008 NZ ETA2, $n = 19$), with SDs ranging from 4.79 (2009 NZ ETA4, $n = 14$) to 7.66 (2009 Au ETA3, $n = 68$). The 2008 Au cohort's mean ETA score remained quite constant throughout the assessments with a very slight increase from ETA1 to ETA4. The 2009 Au cohort's mean ETA scores showed more variation, ranging from 83.96% (ETA 4) to 86.07% (ETA3) and decreasing from ETA1 to ETA4; the 2010 Au cohort's mean scores increased slightly from ETA1 to ETA2. All NZ cohorts' mean ETA scores showed more fluctuation than the corresponding Au mean ETA scores, with no consistent pattern discernible from the data; most NZ mean ETA scores were in the range 84-86%, with the exceptions of 2008 ETA2 (87.73%, n =19) and 2009 ETA3, (79.30%, n =7). See Table 41**.**

Summary of findings – descriptive statistics

This study identified that roughly one in three candidates were selected into RACS GS training in any year. In 2008 and 2009, around 100 trainees were selected into GS training, increasing to 140 in 2010. Approximately four in five of those selected entered the Au program and the rest entered the NZ program. Mean scores in selection were high, particularly in the RR and Int, contributing to high Total sel scores – mean Total sel scores ranged from 73% to 77%. Mean scores in examinations were consistent for the three cohorts and between both countries. Trainees tended to score higher in the written examinations than in the practical, CE, with trainee performance in the GenSSE slightly stronger than in the SpecSSE. Mean scores for all work-based assessments were high. DOPS mean scores ranged from 79% to 86%; with no longitudinal pattern of performance discernible. MiniCEX mean scores had the broadest range of the work-based assessments (80%–92%), with no discernible longitudinal pattern of performance. ETA mean scores (79%–88%) were similar to DOPS mean scores, with consistency between Au and NZ scores.

Pearson correlation analyses within selection and assessment categories (intra-assessement)

Pearson product-moment correlation analyses were conducted to evaluate the degree of associations for trainee performance scores in selection, in examinations, in DOPS, in MiniCEX, and in ETAs. Statistically significant values ($p < .05$ and $p < .01$) are highlighted in correlation tables. In this analysis, values of r are considered from ‘very weak, negligible’ to ‘very strong’ as described in Table 12 *Guide to interpreting correlation coefficients (r)* in the Methods chapter.

Selection items

Table 44 shows that eight significant correlations were found for performance across the three selection instruments (CV, RR, Int). All significant correlations between selection instruments involved the Int, were weak to moderate, and inverse. Performance in the Au CV had a weak to moderate, inverse relationship with performance in the Au Int for two cohorts (2009 and 2010). The only other significant association in any year was a weak, inverse association between Au RR and Au Int, in 2010. Apart from these associations, no single selection tool is closely related to performance in either of the other selection tools. The pattern of associations observed between all selection items and Total sel scores in all cohorts was framed by combining the selection item scores to form the Total sel score.

No significant correlations were found between scores in selection instruments for 2008. In 2009 a moderate, negative association was found between Au CV and Au Int, $r = -.38$, $p < .001$ (2-tailed); this association was maintained in the 2009 ANZ result. A weak negative association was also found between Au CV and Au Int, $r = -.22$, $p < .05$ (2-tailed) in 2010, combined with a weak, negative association between Au RR and Au Int, $r = -.18$, $p < .05$ (2-tailed). When data for the three cohorts was combined, negative associations were identified between Au CV and Au Int, $r = -.26$, $p < .05$ (2-tailed) and between NZ RR and NZ Int, $r = -.38$, $p < .01$ (2-tailed).

Examinations

Thirty-two significant correlations were discerned between performance in the three examinations (GenSSE, SpecSSE, CE), with all Au and NZ cohorts exhibiting very strong associations between performance in the two written examinations (GenSSE and SpecSSE) and moderate to very strong associations between performance in the written examinations and the practical CE. As shown in Table 45 moderate associations were found for the three Au cohorts, between the GenSSE and the CE. The NZ results differed from the Au results, with the NZ 2009 cohort revealing very strong associations between the SSEs and the CE, but the 2008 and 2010 cohorts showing no significant associations between the SSEs and the CE. When the data for the selection years were combined, the moderate to very strong correlations between the examinations for both Au and NZ were confirmed. The GenSSE and SpecSSE are very strongly associated with each other and each of these SSE examinations is moderately correlated with the CE.

Direct Observation of Procedural Skills (DOPS)

Excluding correlations between individual assessments and the average DOPS, eight weak to moderate significant correlations were identified between DOPS assessments. Data for the 2008 cohort's performance in the four DOPS assessments show moderate, significant associations between Au DOPS1 and Au DOPS2, maintained in the combined ANZ data; NZ associations for the 2008 cohort are based on inadequate data and are thus inconclusive. The 2009 cohort maintained the moderate association between Au DOPS1 and Au DOPS2, however no significant associations were identified in the 2010 cohort. When the three cohorts' data were combined, the associations between Au DOPS1 and Au DOPS2 were reflected and new associations were revealed between DOPS2 and DOPS3. Associations were identified between all DOPS assessments and Average DOPS scores in all cohorts. See Table 46 and Table 47.

Mini Clinical Evaluation Exercises (MiniCEX)

Excluding correlations between individual assessments and the average MiniCEX, nine weak to very strong significant correlations were identified between MiniCEX assessments, with most being moderate. No significant associations were identified in MiniCEX performance for the 2008 cohort; in 2009 a strong association was identified between performance in the Au MiniCEX1 and Au MiniCEX3. NZ associations in all cohorts in the later assessments (MiniCEX3 and MiniCEX4) are based on inadequate data and are thus unreliable. The combined cohorts' data confirmed the relationship between Au MiniCEX1 and Au MiniCEX3 and revealed an additional, weak relationship between ANZ MiniCEX2 and ANZ MiniCEX3, and an additional moderate relationship between Au MiniCEX3 and Au MiniCEX4. Associations were identified between all MiniCEX assessments and Average MiniCEX scores in all cohorts. See Table 48 and Table 49.

End of term assessments (ETAs)

Excluding correlations between individual assessments and the average ETA, twenty-one significant correlations were identified for performance in the four ETA assessments; all except two of these correlations were weak to moderate – the correlations between the 2009 NZ ETA1 and NZ ETA2, and between the 2009 NZ ETA2 and NZ ETA4 being strong. The 2008 cohort demonstrated a moderate, significant association between NZ ETA2 and NZ ETA4; the combined ANZ data demonstrated a weak, significant association between ETA3 and ETA4, indicating that the Au and possibly the NZ associations between these assessments approached significance. The 2009 cohort presented two key associations for ETA performance—between ETA1 and ETA2 (weak for Au and strong for NZ) and between ETA2 and ETA4 (moderate for Au and strong for NZ). No significant relationships were identified for the 2010 cohort due to limited data. When data for the three cohorts were combined, associations were found between performance in all ETA assessments; these were not uniform across the two countries. End of Term Assessments (ETA)

Table 50 and Table 51 provide ETA correlation information. Associations were identified between all ETA assessments and Average ETA scores in all cohorts.

Summary of findings – Pearson correlation analyses within selection and assessment categories

Eight significant correlations were identified for performance across the three selection instruments. The only significant correlations discerned were between the Int and the CV and the Int and the RR; all significant correlations for selection were negative and weak to moderate. Performance in the three examinations was highly consistent, with 32, moderate to very strong significant correlations identified. Excluding correlations between individual DOPS and MiniCEX assessments and the Average DOPS or Average MiniCEX scores, eight (four weak and four moderate) significant correlations were identified within DOPS and nine weak to very strong significant correlations were identified within MiniCEX assessments, (most being moderate). Most significant correlations in DOPS assessments were identified between DOPS1and DOPS2. Moderate, significant correlations among MiniCEX assessments in combined years were identified between MiniCEX1 and MiniCEX3, and between MiniCEX3 and MiniCEX4, and a weak, significant correlation was identified between MiniCEX2 and MiniCEX3. Excluding correlations between individual assessments and the average ETA, twenty-one significant correlations were identified within performance in ETAs, spanning most

ETA assessments. Most were weak to moderate. The combined years' data reveal associations between performance in all ETA assessments, with NZ associations being stronger than those identified for Au.

Part two: Comparison of trainee performance in selection and in subsequent assessments

Part two of the study compared trainee performance in selection with performance in subsequent assessments. Pearson correlations and regression analyses were conducted to determine the degree of association between performance in the selection items and performance in each of the subsequent assessment items, (i.e. selection vs. assessment).

Pearson correlation analyses comparing selection with assessments: summary results

Number and strength of associations

Pearson correlations were used to identify associations between scores in selection and scores in assessments where n > 5. Results are presented in Table 14, Table 15 and Table 16 and in Figure 12 to Figure 16. From the available data, 776 associations between scores in selection and scores in assessments were possible, representing Examinations (n = 144), DOPS (n = 208), MiniCEX (n = 208) and ETA (n = 216). See Correlations possible and identified

Table 56. Of these possible associations, 186 significant associations were identified for n > 5, a ratio of 1:4.17 actual to possible associations. See Note: - = no data available, or n ≤ 5

Table 57. Four additional significant associations were identified for n ≤ 5. See Table 63. Data were available for 16 fewer NZ assessments, each of which could have 4 possible associations with the selection items, resulting in 64 fewer possible significant associations for NZ (216 possible associations) than for Au (280 possible associations). DOPS3, DOPS4, MiniCEX3 and MiniCEX4 were the assessments where there were fewer than five NZ trainee results.

Table 14 *Nominal strength of Pearson correlation coefficients (r) and number and percentage of total identified*

	Coefficient, *r*		Total identified	
Strength of association	Positive	Negative/Inverse	n in this range (% of 186)	
Very weak, negligible	.01 to .15	-.01 to -.15	19	(10.22%)
Weak	.16 to .29	-.16 to -.29	100	(53.76%)
Moderate	.30 to .49	-.30 to -.49	40	(21.51%)
Strong	.50 to .69	-.50 to -.69	23	(12.37%)
Very strong	.70 to 1.0	-.70 to 1.0	4	(2.15%)
TOTAL			**186**	

Fifty-four of the 56 significant correlations between RRs and assessment items were positive, as were 41of the 47 significant correlations between Int and assessment items; 36 of the 40 CV significant correlations were negative; the exceptions were the 2010 Au and 2010 ANZ CV, which was positively correlated with DOPS1 and the Combined years' NZ CV, which was positively correlated with the GenSSE. Table 14 presents criteria used to classify the strength of associations, the number of significant correlations (for n > 5) and their percentage of the total significant correlations identified in each range. Figure 12 presents positive and negative correlations identified in each range. Figure 26 presents the number of positive and negative correlations identified per selection item.

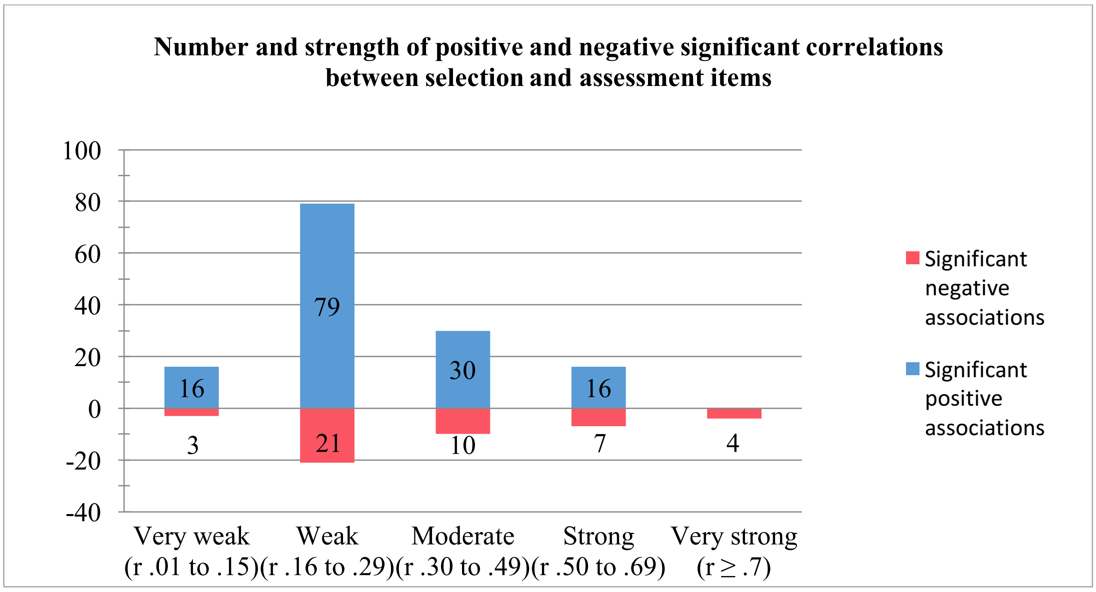

Figure 12. Number and strength of positive and negative significant correlations between selection and assessment items

Significant correlations between selection items and assessment items

One hundred and eighty-six significant correlations were identified between performance in selection items and assessments; 45 correlations were negative, 141 were positive, with most being weak to moderate, as shown in Figure 12. All very weak correlations were identified in the Au and ANZ groups – the ANZ results reflecting the predominance of Au trainees. Most significant NZ associations were moderate to very strong. See Figure 12 and Figure 22.

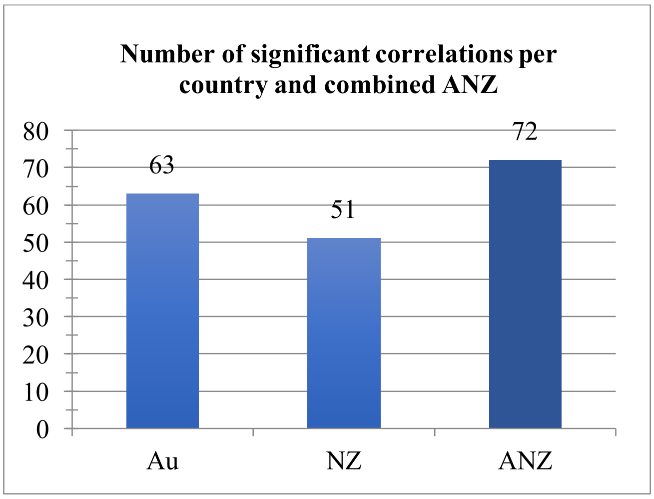

Figure 13. Significant Pearson correlations between selection items and assessment items per country and combined ANZ

When considering the data per country, Au was shown to have more significant correlations than did NZ. Most significant correlations (n = 72) were identified when Au and NZ data were combined. See Figure 13 and Table 15. Of the three yearly cohorts, 2008 showed the fewest significant correlations between selection and assessment performance and 2009 showed the most. See Figure 14 and Table 15; 2010 data was incomplete for some of the clinical assessment items; however, 2010 showed more significant correlations (n = 33) than did 2008 (n = 28).

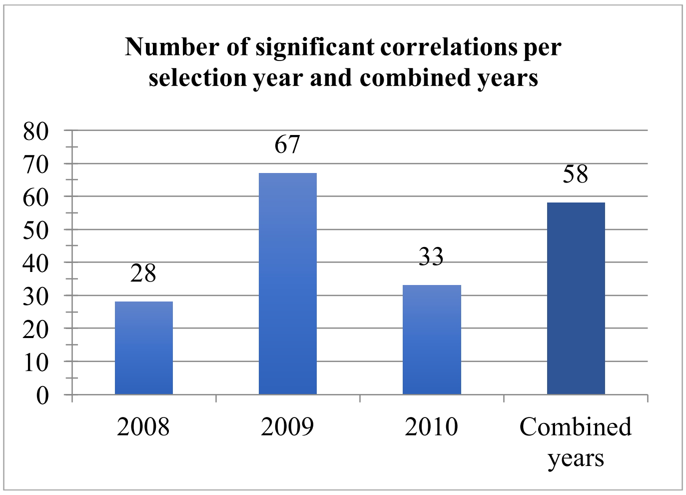

Figure 14. Significant Pearson correlations between selection items and assessment items per year and combined years

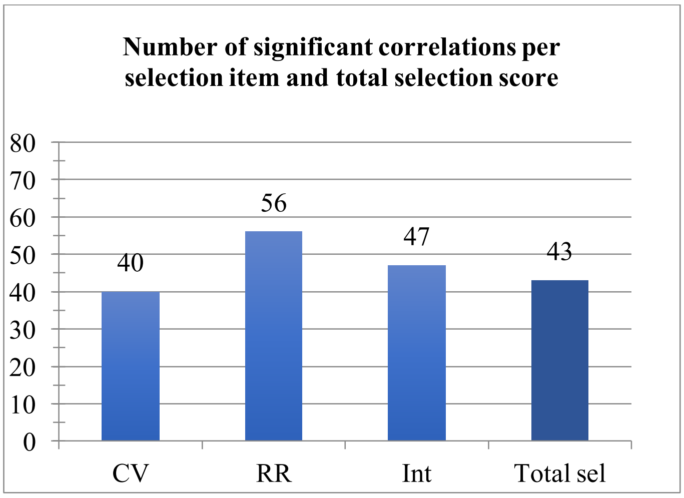

Figure 15. Significant Pearson correlations per selection score and total selection score

When considering correlations per selection item, the RR exhibited the greatest number of significant correlations with subsequent assessment performance (n = 56), followed by the Int (n = 47) and Total sel (n = 43). The CV showed the fewest correlations (n = 40). See Figure 15. When considering correlations per assessment item, two examinations (SpecSSE and CE) each had more than 10 correlations with selection items, with the CE showing more correlations than any other assessment item (n = 19). Of the work-based assessments, all ETAs had 10 or more significant correlations, as did DOPS1 and DOPS2. MiniCEX assessments had the fewest correlations with selection items. See Figure 16.

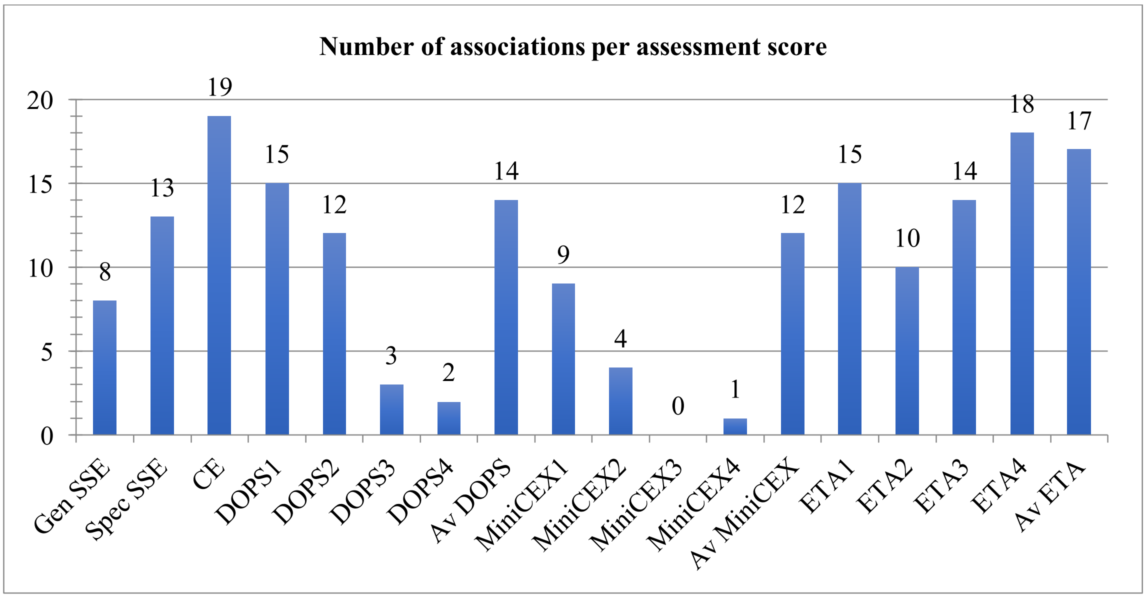

Figure 16. Significant Pearson correlations per assessment score

Very weak significant Pearson correlations.

Sixteen of the nineteen very weak significant correlations were identified when the three annual cohorts' data were combined; seven of the very weak correlations represented Au performance and twelve were ANZ, primarily reflecting the preponderance of Au trainees. The three negative correlations identified were for CV scores. Performance in the three selection items and Total sel was very weakly associated with DOPS, MiniCEX and ETAs and with one examination. See Figure 17 and Table 58.

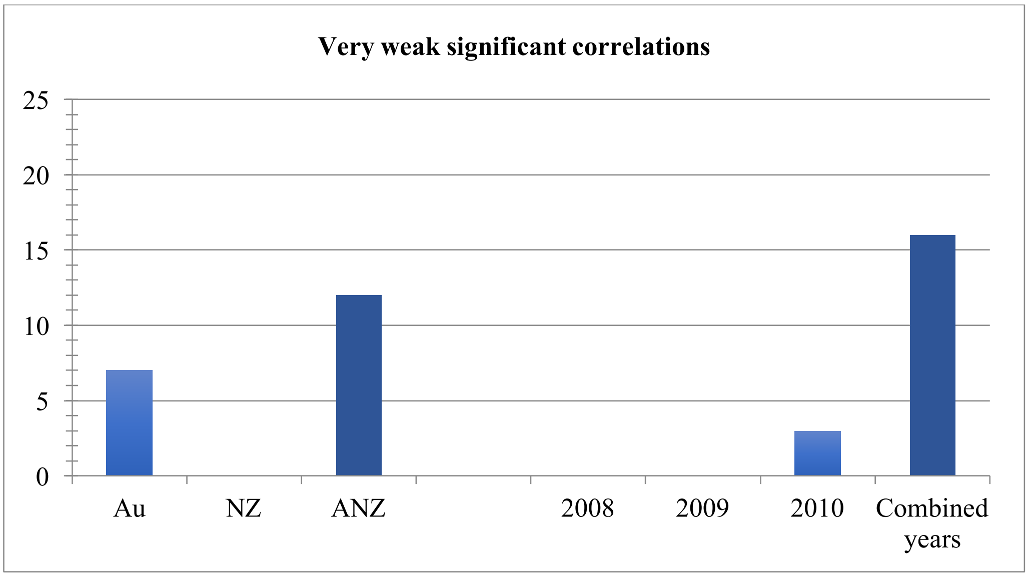

Figure 17. Summary of very weak significant Pearson correlations by country and year

Weak Pearson correlations.

One hundred weak significant correlations were identified, spanning the three annual cohorts; representing Au and ANZ only – no uniquely NZ weak associations were found in any single annual cohort, however nine NZ weak associations were identified when the three annual cohorts' data were combined. Performance in all three selection items and Total sel was weakly associated with DOPS, MiniCEX and ETAs and with all examinations. Seventy-nine of the weak significant correlations were positive. All CV associations were negative except that with the 2010 DOPS1. See Figure 18 and Table 59.

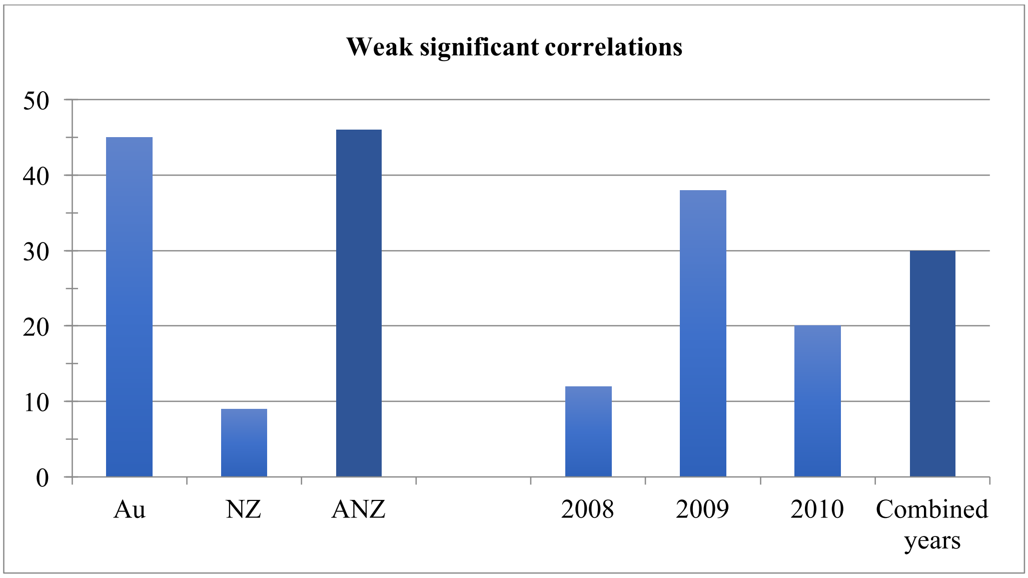

Figure 18. Summary of weak significant Pearson correlations by country and year

Moderate Pearson correlations.

Forty moderate significant correlations were identified (10 negative and 30 positive), predominantly from the 2009 cohort. Seventeen NZ moderate associations were identified, eight of which were detected when the three cohorts' performance data was combined. Performance in all three selection items and with Total sel moderately correlated with DOPS, MiniCEX, ETAs and with all examinations. All CV associations were negative. See Figure 19 and Table 60.

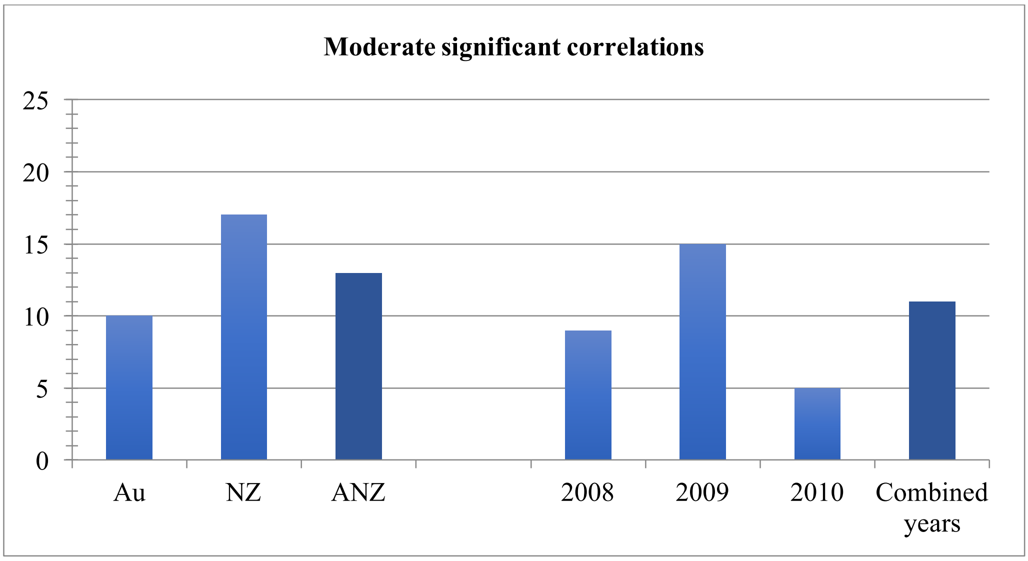

Figure 19. Summary of moderate significant Pearson correlations by country and year

Strong Pearson correlations.

Twenty-three strong significant correlations were identified (7 negative, 16 positive), 21 of these correlations were for NZ. Performance in the three selection items and Total sel was strongly associated with DOPS, MiniCEX and ETAs; the Int and Total sel were also associated with examinations. All CV strong associations were negative. See Figure 20 and Table 61.

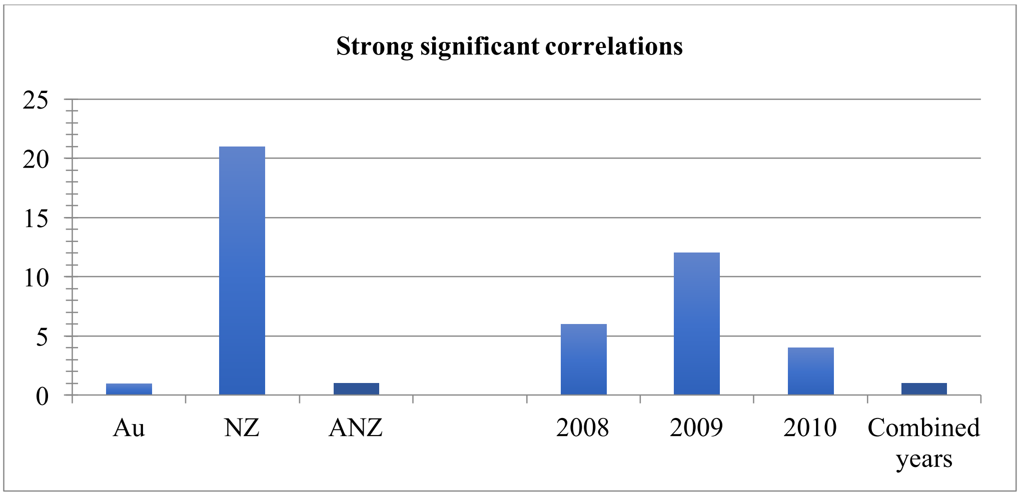

Figure 20. Summary of strong significant Pearson correlations by country and year

Very strong Pearson correlations.

Four very strong significant correlations were all from NZ and were all negative, identifying inverse relationships of the CV and RR with DOPS1 and DOPS2, and between the Int and ETA3 as shown in

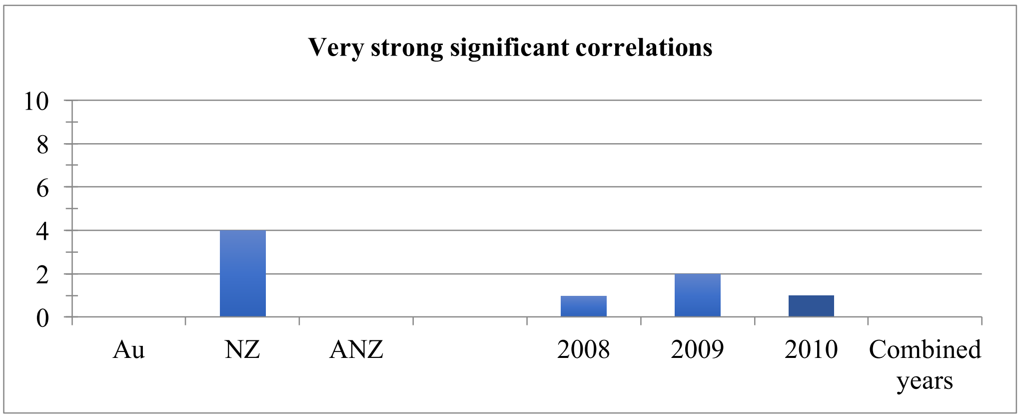

Figure 21 and Table 62.

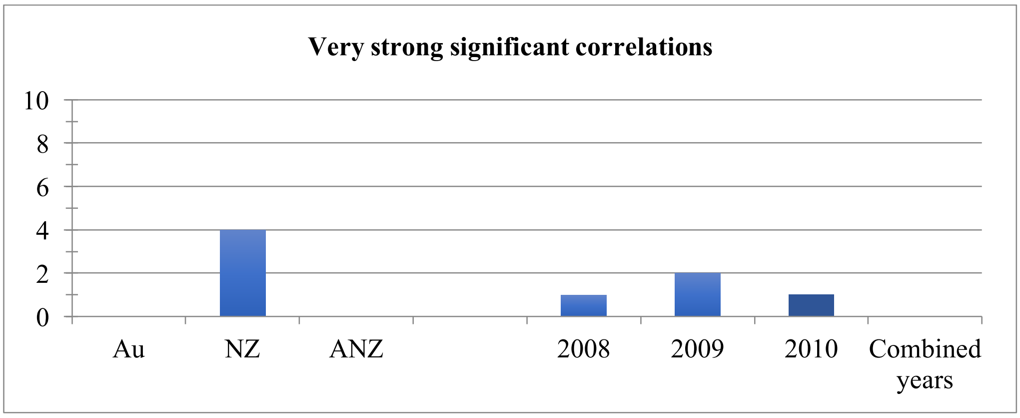

Figure 21. Summary of very strong significant Pearson correlations by country and year

Number and strength of significant correlations per country and per year.

Of the 63 significant correlations identified between Au selection and assessments, 52 were very weak or weak; of the 51 NZ significant correlations between selection and assessments, 28 were strong or very strong. Seventy-two significant correlations were identified for combined ANZ, reflecting the predominant influence of Au in the distribution of very weak to strong associations. See Figure 22.

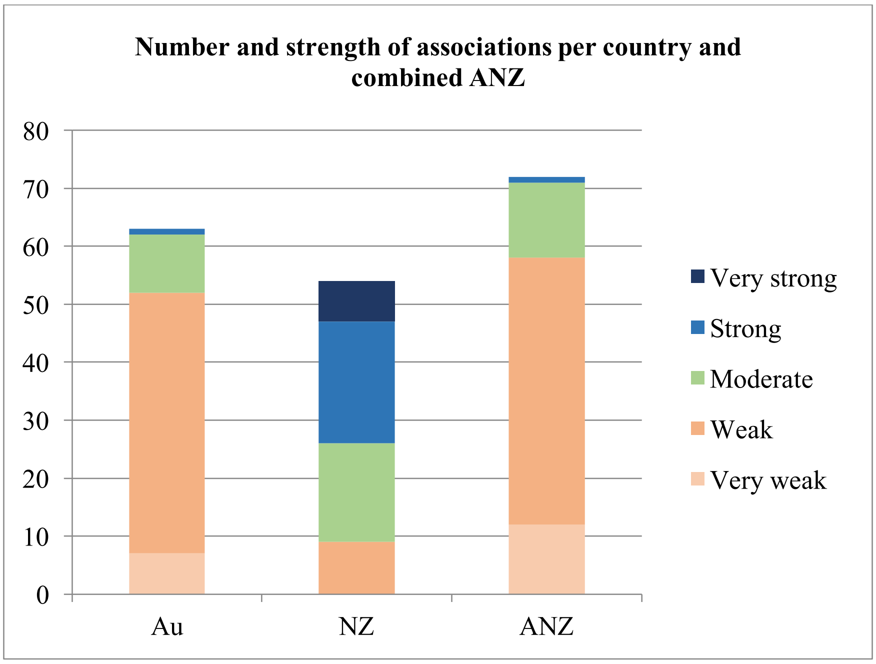

Figure 22. Number and strength of significant Pearson correlations per country and combined ANZ

Of the three yearly cohorts, 2008 had the fewest significant correlations between selection and assessment performance and 2009 had the most; 2010 data was incomplete for some of the clinical assessment performance items; however, 2010 showed more associations than did 2008. The greatest diversity in strength of correlations, ranging from very weak to very strong, was identified in the 2010 cohort. In all cohorts, the largest proportion of correlations were weak; the combined years' data showed the greatest variation of strength of association, spanning the range from very weak to strong. See Figure 23.

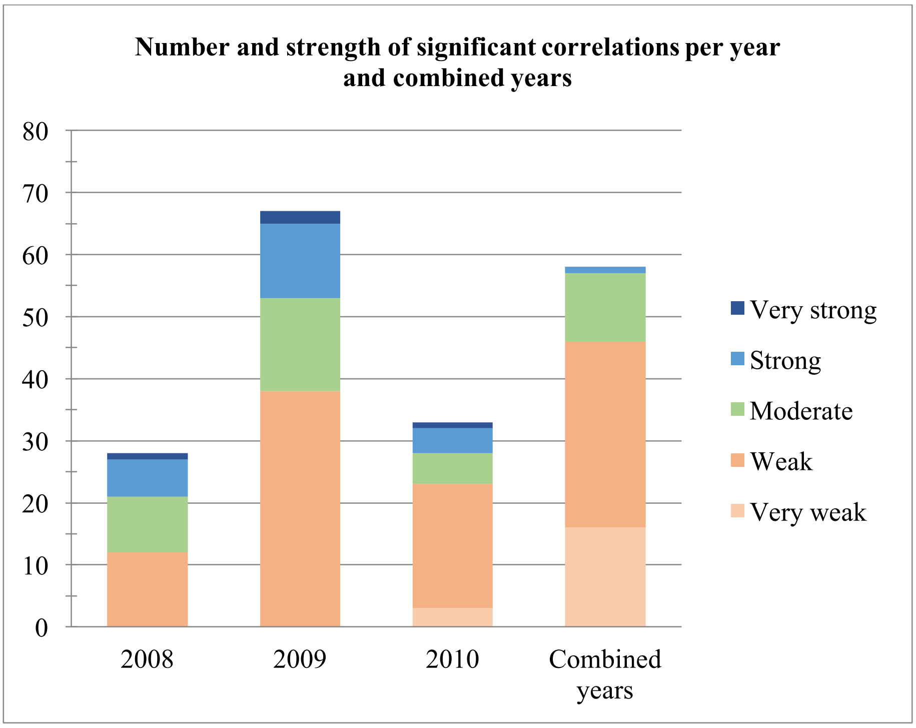

Figure 23. Number and strength of significant Pearson correlations per year and combined years

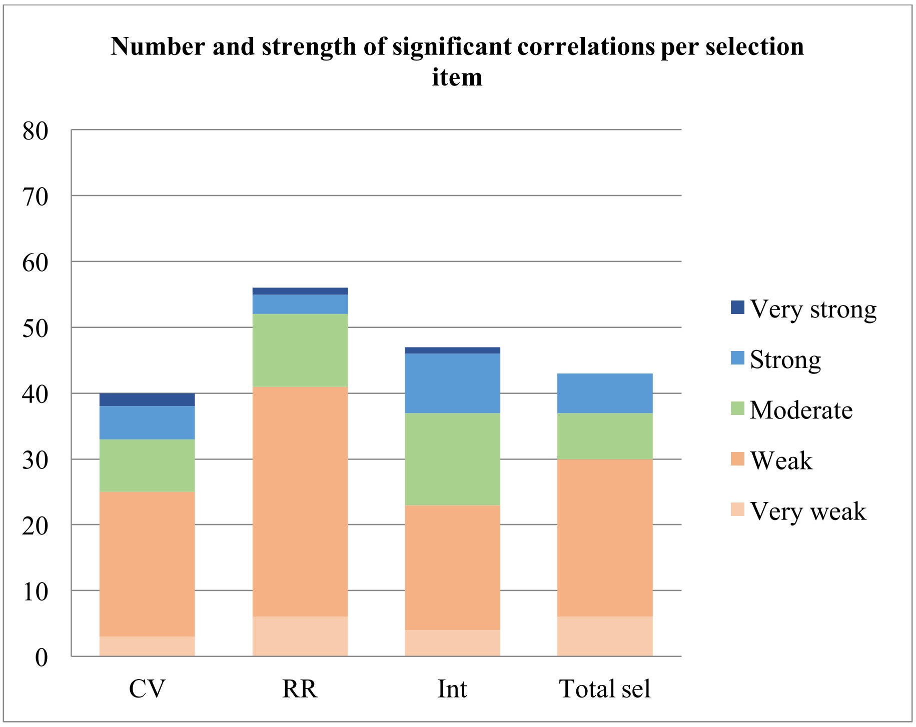

Figure 24. Number and strength of significant Pearson correlations per selection item

Number of significant correlations per selection item by year.

Per selection item, most significant associations were weak. No associations for Total sel were very strong. See Figure 24. On a yearly basis, significant associations between the CV and assessments decreased from 13 (2008) to 4 (2010). Significant correlations between the RR and assessments did not follow a discernible pattern, although in both 2009 and 2010 the RR showed most significant associations of the selection items. For each cohort, the number of significant correlations for the Int lay between the CV and the RR. See Table 15 and Figure 25.

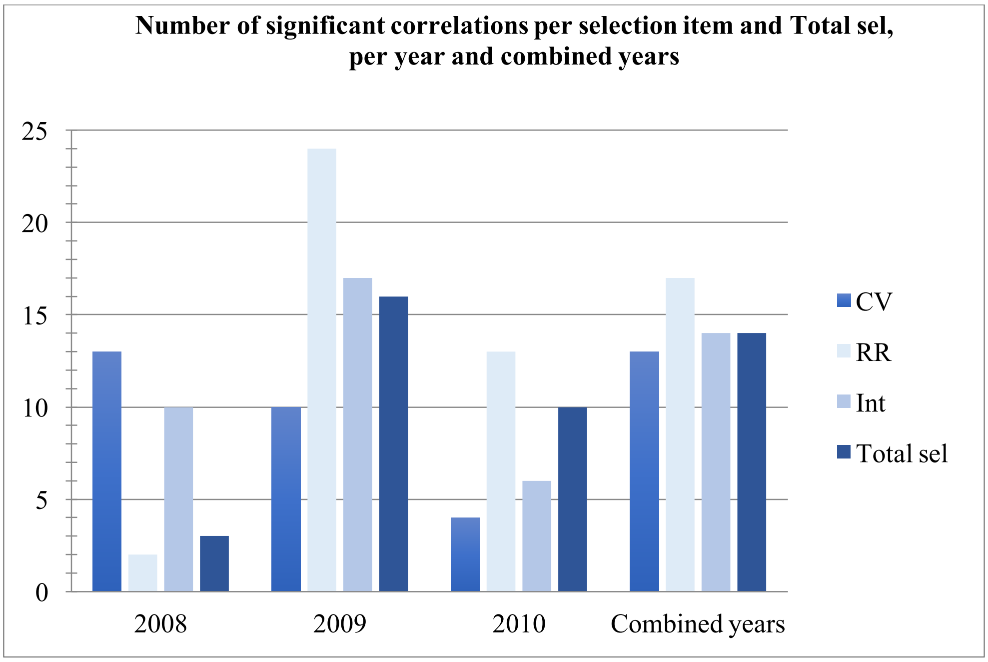

Figure 25. Number of significant Pearson correlations per selection item and Total sel per year and combined years

Number of significant correlations per selection item and country by year.

When analysed by performance across the cohorts, the Au CV and the NZ CV were each significantly correlated with 12 assessments; almost four times as many significant associations were identified for the Au RR than for the NZ RR and almost twice as many significant associations were identified for the NZ Int than for the Au Int; more significant associations were identified for Au Total sel than for NZ Total sel. See Table 15 and Table 16.

Table 15 *Number of significant Pearson correlations per selection item and Total sel; and per country and combined ANZ*

	2008	2009	2010	Combined years	Total	Au	NZ	ANZ	Total
CV	13	10	4	13	**40**	12	12	16	**40**
RR	2	24	13	17	**56**	26	7	23	**56**
Int	10	17	6	14	**47**	11	21	15	**47**
Total sel	3	16	10	14	**43**	14	11	18	**43**
	28	**67**	**33**	**58**	**186**	**63**	**51**	**72**	**186**

Table 16 *Number of significant Pearson correlations for selection items and Total sel, per country and year*

	2008	2009	2010	Combined years	Total	% of 186
	n (%)	**n** (%)	**n** (%)	**n** (%)	**n**	**(%)**
Au	**7**	**23**	**12**	**21**	**63**	**(33.87)**
CV	3	4	1	4	**12**	(6.45)
RR	1	10	6	9	**26**	(13.98)
Int	2	4	2	3	**11**	(5.91)
Total sel	1	5	3	5	**14**	(7.53)
NZ	**10**	**15**	**7**	**19**	**51**	**(27.42)**
CV	5	0	2	5	**12**	(6.45)
RR	0	5	1	1	**7**	(3.76)
Int	4	7	2	8	**21**	(11.29)
Total sel	1	3	2	5	**11**	(5.91)
ANZ	**11**	**29**	**14**	**18**	**72**	**(38.71)**
CV	5	6	1	4	**16**	(8.60)
RR	1	9	6	7	**23**	(12.37)
Int	4	6	2	3	**15**	(8.06)
Total sel	1	8	5	4	**18**	(9.68)
Total	**28** (15.05)	**67** (36.02)	**33** (17.74)	**58** (31.18)	**186**	

Significant positive and negative correlations per selection item and country by year.

One hundred and forty-one significant positive correlations and 45 significant negative correlations and were detected. Most positive correlations were identified for the RR (n = 54).

The Int (n = 41) and Total sel (n = 42) were alike apropos positive correlations. The CV correlations were predominantly negative (n = 36), with very few significant positive correlations (n = 4). When considering the yearly cohorts, most positive correlations were identified for the 2009 cohort (n = 54), noting that the 2010 data were incomplete. Combining the data for all years revealed more correlations than for either 2008 or 2010. When considering the data by country, combining both countries' data revealed more significant positive correlations (n = 72) than for either country individually (Au n = 63; NZ n = 71).

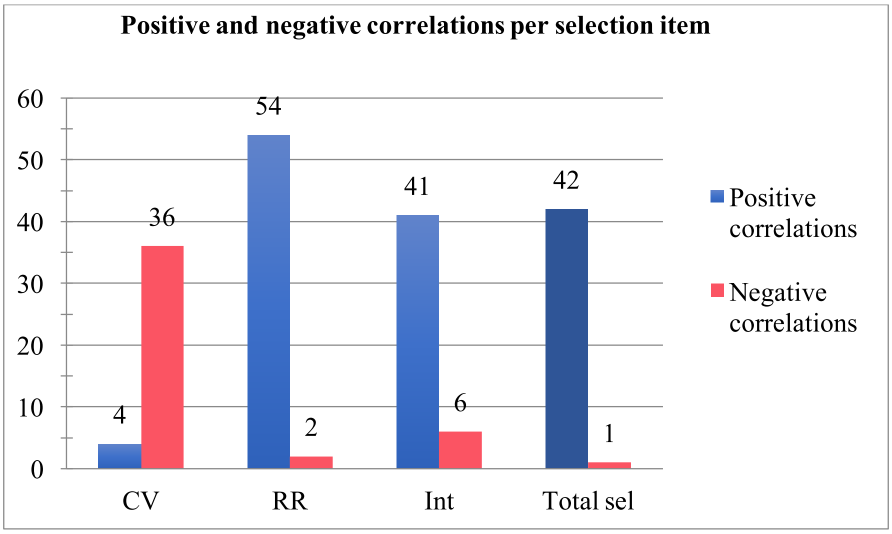

Figure 26. Number of significant positive and negative Pearson correlations per selection item

Most negative correlations were identified for the 2008 cohort (n = 16), being the only year in which negative correlations outnumbered positive correlations. Negative correlations were distributed uniformly between the two countries. See Figure 27 and Figure 28.

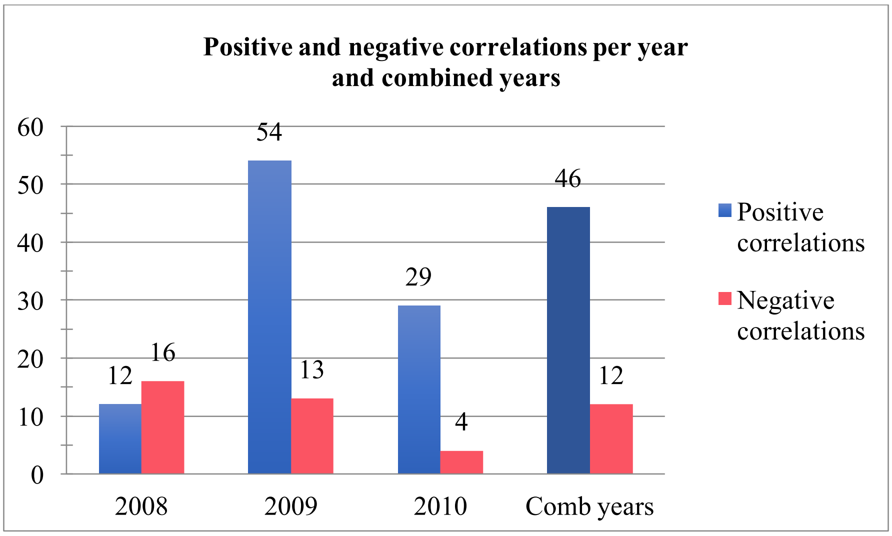

Figure 27. Number of significant positive and negative Pearson correlations per year and combined years

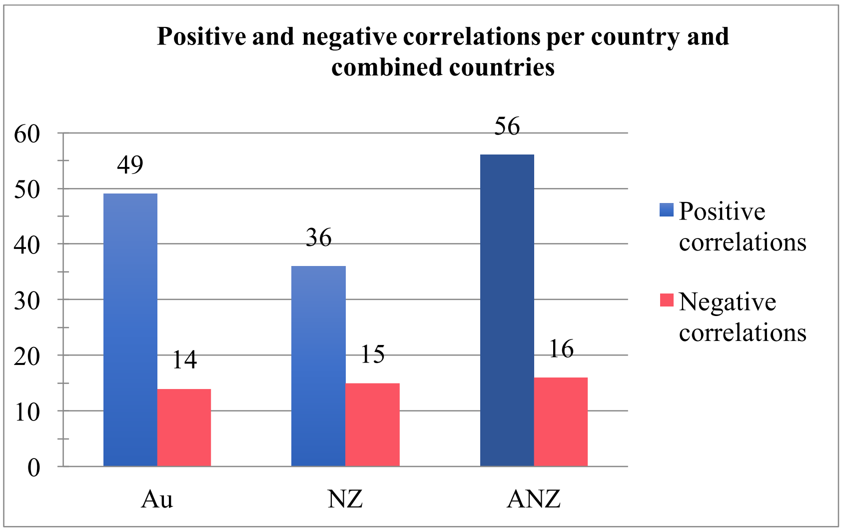

Figure 28. Number of significant positive and negative Pearson correlations per country and combined countries

The CE, ETA1 and ETA4 had the more positive correlations with selection items than did other assessments. DOPS2 and DOPS3 were the only assessments in which negative correlations outnumbered positive correlations. No negative correlations were identified for the GenSSE, MiniCEX3 and MiniCEX4 or ETA1. See Figure 29.

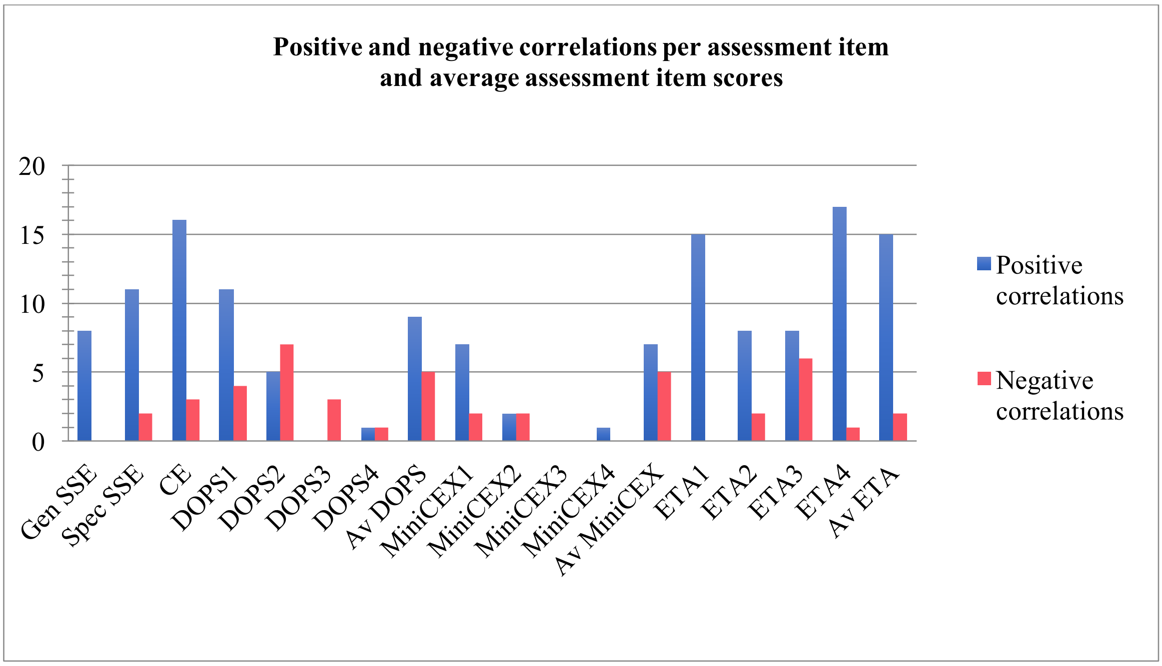

Figure 29. Number of significant positive and negative Pearson correlations per assessment item

Pearson correlation analyses comparing selection with assessments: detailed results

Pearson correlations were determined to establish associations between performance in selection and performances in examinations, DOPS, MiniCEX and ETA assessments.

Correlations between GS selection and examination performance

Variable relationships were found between selection performance and examination performance. No associations were found between selection and examination performance for the 2008 cohort. In the subsequent two cohorts, descriptives indicated weak to moderate, negative associations between the CV and examination performance, a single, moderate association between the RR and examination performance and weak to strong associations between the Int and examination performance. Associations between Australian selection items and assessments tended to be weak, while NZ associations were mostly moderate to strong. The 2009 cohort showed most associations for both countries; the 2009 Int was associated with all three examinations, and all 2009 Au selection items were associated with performance in the CE (the CV negatively so).

The 2009 Au CV and ANZ CV scores showed weak, negative associations with performance in the SpecSSE and the CE. The 2009 Au RR had a weak association with the CE; ANZ RR scores approached a significant correlation with the CE. Both the 2009 Au Int and 2009 NZ Int had weak to strong associations with performance in all three examinations , with the NZ associations being stronger than those for Au and associations tending to be stronger for the CE than for the written examinations. All Au 2009 selection items had weak to moderate association with CE performance; the CV association being negative, all others positive.

The 2010 NZ CV scores showed moderate, negative association with performance in the CE. Weak to strong associations were identified between the Int and all three examinations—the Au Int being weakly associated with the CE, and the NZ Int being strongly associated with the two written examinations. The NZ and Au Total sel scores were similarly associated with all three examinations.

The combined years' data reflected the associations identified for individual yearly cohorts; the combined years' associations were predominantly weak for the CE and moderate to strong for the SSEs; the NZ selection being associated with the two SSEs and Au selection being predominantly associated with the CE. All selection instruments were associated with one or more examinations; the Int and Total sel scores were associated with all examinations, although there were differences between the two countries. A weak association of the NZ CV with the GenSSE was not identified in any individual year and is the only instance where the CV has a significant, positive association with examination performance. The Au RR, Au Int and Au Total sel associations with the CE reflected the associations of these selection items in 2009 and 2010. Little consistency was identified between the two countries' selection-to-examination correlation outcomes. See Table 17.

Table 17 *Pearson correlations (r) between GS ANZ selection and examination performance for 2008, 2009, 2010 selection years and combined selection years*

		Pearson Correlations (*r*)								
		GenSSE %			SpecSSE %			CE %		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
2008										
	CV %	-.11	.31	-.04	-.12	.36	-.07	.00	.26	.04
	RR %	.00	-.08	-.01	.11	.16	.12	.16	-.18	.12
	Int %	-.03	.22	-.01	.04	-.22	.04	.07	.13	.08
	Total sel %	-.07	.32	-.03	.05	.33	.07	.14	.13	.14
N	100	78	18	96	56	10	66	79	18	97
2009										
	CV %	-.10	.21	-.04	**-.23***	.03	**-.21***	**-.22***	.01	**-.18***
	RR %	.10	-.35	.05	.13	-.01	.11	**.19****	-.09	.16
	Int %	**.19***	**.43***	**.20***	**.21***	**.67****	**.26****	**.39****	**.46***	**.38****
	Total sel %	.14	.18	.15	.09	.41	.12	**.28****	.12	**.27****
N	107	88	17	105	80	13	93	88	17	105
2010										
	CV %	-.04	.21	-.01	.02	.02	.04	.01	**-.48***	-.05
	RR %	.08	.35	.13	.06	.21	.10	.13	.24	.12
	Int %	-.03	**.68****	-.04	.09	**.69****	.07	**.18***	.23	**.19****
	Total sel %	.00	**.69****	.05	.12	**.54****	**.14***	**.22****	.04	**.19****
N	140	118	19	137	114	18	132	116	19	135
Combined years										
	CV %	-.08	**.25***	-.04	-.09	.12	-.06	-.08	-.06	-.08
	RR %	.01	.09	.03	.03	**.28***	.07	**.22****	.12	**.21****
	Int %	.02	**.32****	.03	**.10***	**.29***	**.11***	**.20****	.20	**.20****
	Total sel %	-.03	**.44****	.02	.03	**.48****	.08	**.24****	.20	**.24****
N	347	284	54	338	250	41	291	283	54	337

Notes. CV% = CV score as a percentage. RR% = Referee reports score as a percentage. Int % = Interview score as a percentage. Total sel % = Total selection score as a percentage. CE% = Clinical Exam score. GenSSE % = Generic Surgical Sciences Exam score. SpecSSE% = Specialty-specific Surgical Sciences Exam score.
r = Pearson Correlation. N = Number of trainees.
** = Correlation is significant at .01 (1-tailed). * = Correlation is significant at .05 (1-tailed). Teal = Also significant in regression analysis. Diagonal = Pearson correlation is not significant at .05 but is significant in regression analysis.

Correlations between GS selection and Direct Observation of Procedural Skills (DOPS)

Pearson's correlation coefficients between performance in four selection items (including the Total sel) and four Direct Observation of Procedural Skills (DOPS) Assessments (DOPS1, DOPS2, DOPS3, DOPS4) and for the average DOPS scores were calculated. See Table 18.

As identified with the examination results, associations between Au selection items and DOPS assessments tended to be weak, while NZ associations were moderate to strong. There were two instances of the same Au and NZ selection instruments being significantly associated with particular DOPS assessments—the 2008 Au CV and NZ CV with DOPS1 and combined years' CVs with DOPS2.

The 2008 cohort's, Au CV showed a weak, negative association with DOPS1 while the NZ CV was strongly, negatively associated with both DOPS1 and DOPS2. There was inadequate data to calculate associations between NZ selection and DOPS3 or DOPS4. The associations of the Au and NZ CVs with DOPS1 and DOPS2 were corroborated by Au and NZ CV associations with the average DOPS. The 2008 NZ Int had a strong association with DOPS1 and NZ Total sel had a strong association with DOPS2. Overall, the 2008 NZ associations with DOPS assessments were more numerous and stronger than those for Au. All 2008 CV associations with DOPS performance were negative.

The 2009 NZ RR was very strongly, negatively associated with DOPS1, the Au RR was weakly associated with DOPS2, the 2009 NZ Int was strongly associated with DOPS2 and the 2009 Au Int was strongly, negatively associated with DOPS4. The Au Total sel score was weakly associated with the Average DOPS score, although not with any individual DOPS assessments. As identified in the 2008 cohort, there was inadequate data for NZ associations with DOPS3 and DOPS4 to be reliable. There was no synchronicity between 2009 Au and NZ results.

The 2010 Au CV, Au RR and Au Total sel scores showed weak association with performance in DOPS1 and the Au Int showed a moderate, negative association with performance in DOPS3. The Au RR association with DOPS1 was maintained in the Average DOPS, indicating that Au RR associations with subsequent DOPS assessments were approaching significance. The only significant association between NZ 2010 selection and DOPS performance was a very strong, negative association of the NZ CV with DOPS2. Au selection performance was not associated with DOPS4 and there was no NZ data available for DOPS3 or DOPS4.

The combined years' data confirmed the relationships identified in the 2010 cohort of Au RR, Au Total sel and ANZ Total sel with DOPS1 and highlighted negative associations of the Au Int, Au CV and NZ CV with DOPS2, (this moderate, negative association between the NZ CV and DOPS2 particularly reflected the 2008 and 2010 cohorts). No significant associations were observed between selection performance and DOPS3. A weak Au RR relationship with DOPS4 was identified. All selection instruments were associated with one or more DOPS assessments. As identified in individual cohorts, Au associations were weak and NZ associations were predominantly strong.

Table 18 *Correlations between GS ANZ selection performance and DOPS performance for 2008, 2009, 2010 and combined selection years; n ≥ 5*

		Pearson Correlations (*r*)														
		DOPS1 %			DOPS2 %			DOPS3 %			DOPS4 %			Average DOPS %		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
2008																
	CV %	-.25*	-.62**	-.30**	-.23	-.63*	-.28*	-.30	Y	-.36*	.22	Q	.22	-.27*	-.70**	-.32**
	RR %	.03	-.10	-.00	.10	.20	.09	.04	Y	.04	.17	Q	.17	.07	.05	.05
	Int %	.22	.50*	.22*	.15	-.33	.06	-.01	Y	-.00	-.07	Q	-.07	.06	.26	.05
	Total sel %	.08	-.31	.02	.07	-.52*	-.04	-.12	Y	-.15	.19	Q	.19	-.02	-.39	-.08
N	100	45	14	59	39	11	50	20	2	22	10	0	10	45	14	59
2009																
	CV %	-.02	-.11	-.03	-.12	-.19	-.12	.20	Y	.12	-.07	Y	.00	.03	-.09	.02
	RR %	.18	-.73**	.09	.27*	-.10	.25*	.22	Y	.22	.43	Y	.31	.26*	-.60*	.20*
	Int %	.08	.24	.10	.13	.64*	.17	-.12	Y	-.11	-.59*	Y	-.40	.08	.46	.12
	Total sel %	.17	-.18	.11	.20	.39	.22*	.28	Y	.22	-.09	Y	.-.03	.27*	.01	.23*
N	107	64	10	74	61	8	69	30	2	32	13	2	15	66	10	76
2010																
	CV %	.23**	.09	.22**	-.11	-.75*	-.13	-.01	Q	-.01	-.33	Q	-.33	.04	-.16	.03
	RR %	.17*	.11	.15*	.02	.36	.03	.23	Q	.23	.38	Q	.38	.18*	.32	.18*
	Int %	.00	-.26	.01	.09	-.30	.06	-.48**	Q	-.48**	-.02	Q	-.02	.00	-.31	.00
	Total sel %	.29**	-.08	.25**	-.03	-.26	-.05	-.07	Q	-.07	-.04	Q	-.04	.15	-.11	.13
N	140	102	17	119	77	8	85	34	0	34	17	0	17	103	17	120
Combined years																
	CV %	.06	-.23	.03	-.16*	-.48**	-.19**	-.02	Y	-.06	-.14	Y	-.09	-.01	-.35**	-.05
	RR %	.14*	-.30	.08	.12	.02	.10	.14	Y	.13	.27*	Y	.24	.19**	-.21	.13*
	Int %	.09	.18	.09	.12*	.15	.10	-.17	Y	-.16	-.18	Y	-.15	.04	.21	.05
	Total sel %	.19**	-.24	.13*	.06	-.20	.02	-.03	Y	-.06	.00	Y	.04	.14*	-.21	.09
N	347	211	41	252	**177**	27	204	84	4	88	40	2	42	214	41	255

Notes. CV% = CV score as a percentage. RR% = Referee reports score as a percentage. Int % = Interview score as a percentage. Total sel % = Total selection score as a percentage. DOPS % = DOPS Assessment score. Average DOPS % = Mean score of DOPS 1-4.
r = Pearson Correlation. N = Number of trainees. Y = no scores listed because n < 5 (see Table 64). Q = No results because n = 0 (see Table 64)
** = Correlation is significant at .01 (1-tailed). * = Correlation is significant at .05 (1-tailed). Teal = Also significant in regression analysis.

Correlations between GS selection and Mini Clinical Evaluation Exercise (MiniCEX)

Pearson's correlation coefficients between performance in four selection items (including the Total sel) and four Mini Clinical Evaluation Exercise Assessments: MiniCEX1, MiniCEX2, MiniCEX3, MiniCEX4 and for average MiniCEX scores were calculated. See Table 19.

As seen with results for the examinations and DOPS, significant associations between Australian selection items and MiniCEX assessments tended to be weaker than NZ associations. Au and NZ selection tools were not aligned in their associations with MiniCEX assessments—there were no instances of both countries having the same tool significantly associated with a particular MiniCEX assessment.

The 2008 cohort's Au Int was moderately associated with MiniCEX1 and the NZ CV was strongly, negatively associated with MiniCEX2 and with Average MiniCEX. The 2009 Au CV was weakly, negatively associated with MiniCEX1 and with Average MiniCEX; the 2009 Au RR was weakly associated with MiniCEX2 and with Average MiniCEX. All these associations were also identified in ANZ associations. The 2010 Au RR was weakly associated with MiniCEX1, the association being maintained for ANZ RR and for Average MiniCEX. No associations were detected with later MiniCEX assessments.

The combined years' data confirmed the negative association between NZ CV and MiniCEX2 and revealed a moderate relationship between the NZ Int and MiniCEX1. The combined years' data confirmed the weak relationship between Au RR and the Average MiniCEX and the negative relationship between NZ CV and Average MiniCEX; the relationship between NZ Int and MiniCEX1 was maintained in the Average MiniCEX.

All selection instruments were associated with one or more MiniCEX assessments; fewer associations were identified for MiniCEX than for DOPS assessments. Most significant associations were weak to moderate.

Table 19 *Correlations between GS ANZ selection and MINICEX performance for 2008, 2009, 2010 selection years and combined selection years; n ≥ 5*

		Pearson Correlations (*r*)														
		MiniCEX1 %			MiniCEX 2 %			MiniCEX 3 %			MiniCEX 4 %			Average MiniCEX %		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
2008																
	CV %	-.11	-.38	-.18	-.03	**-.58***	-.15	-.17	Y	-.29	-.04	Q	-.04	-.07	**-.60****	-.20
	RR %	-.00	.09	.01	.18	.19	.18	-.05	Y	-.04	.27	Q	.27	.07	.23	.09
	Int %	**.30***	.43	**.31****	-.01	-.22	-.03	.05	Y	.07	-.13	Q	-.13	.05	.23	.06
	Total sel %	.16	.01	.14	.09	-.44	.02	-.07	Y	-.12	.11	Q	.11	.05	-.16	.01
N	100	46	14	60	41	10	51	20	2	22	9	0	9	47	14	61
2009																
	CV %	**-.22***	-.06	**-.19***	-.14	-.43	-.16	.01	Y	-.06	-.49	Y	**-.50***	**-.27****	-.13	**-.24***
	RR %	.13	-.21	.12	**.25***	-.31	**.23***	.29	Y	.29	.23	Y	.23	**.26***	-.28	**.25***
	Int %	.09	.32	.10	.05	.54	.06	.04	Y	.03	.16	Y	.11	.14	.46	.14
	Total sel %	.00	.09	.02	.10	.05	.09	.29	Y	.22	-.21	Y	-.24	.10	.11	.11
N	107	67	10	77	58	8	66	24	2	26	10	2	12	68	10	78
2010																
	CV %	.02	.15	.02	-.01	.13	.00	-.00	Q	-.00	-.12	Q	-.12	-.01	.22	.00
	RR %	**.21***	.36	**.21****	.09	.22	.11	.09	Q	.09	-.01	Q	-.01	**.18***	.26	**.19***
	Int %	-.03	-.00	.00	-.02	.25	-.08	.04	Q	.04	-.05	Q	-.05	-.02	.11	-.03
	Total sel %	.13	.23	**.15***	.03	.29	.02	.08	Q	.08	-.13	Q	-.13	.10	.29	.11
N	140	100	17	117	74	8	82	34	0	34	12	0	12	101	17	118
Combined years																
	CV %	-.07	-.22	-.09	-.09	**-.39***	-.11	-.02	Y	-.08	-.22	Y	-.24	-.10	**-.33***	**-.12***
	RR %	.10	-.01	.09	.09	.22	.11	.05	Y	.05	.21	Y	.21	**.13***	.05	**.13***
	Int %	.10	**.31***	**.12***	.02	.07	-.01	.05	Y	.06	-.03	Y	-.04	.05	**.31***	.05
	Total sel %	.09	.09	.08	.01	-.03	-.01	.06	Y	.02	.01	Y	-.01	.05	.07	.05
N	347	213	41	254	173	26	199	78	4	82	31	2	33	216	41	257

Notes. CV% = CV score as a percentage. RR% = Referee reports score as a percentage. Int % = Interview score as a percentage. Total sel % = Total selection score as a percentage. MiniCEX % = MiniCEX Assessment score. Average MiniCEX % = Mean score of MiniCEX 1-4.
r = Pearson Correlation. N = Number of trainees. Y = no scores listed because n < 5 (see Table 65). Q = No results because n = 0 (see Table 65).
** = Correlation is significant at .01 (1-tailed). * = Correlation is significant at .05 (1-tailed). Teal = Also significant in regression analysis.

Correlations between GS selection and End of Term Assessments (ETAs)

Pearson's correlation coefficients between performance in four selection items (including the Total sel) and four End of Term Assessments (ETAs): ETA1, ETA2, ETA3, ETA4 and for the average ETA scores were calculated. See Table 20.

No 2008 selection scores were associated with performance in ETA1. The Au Int was weakly, negatively associated with performance in ETA2 and the NZ Int was moderately associated with ETA3, ETA4 and the Average ETA. The Au CV had a weak, negative association with performance in ETA3. ETA4 had most associations with selection, being associated with Au RR, Au Total Sel and NZ Int. The NZ Int was the only selection item to demonstrate an association with the 2008 cohort's Average ETA score.

The number of associations between selection performance and ETA performance increased as trainees in the 2008 cohort progressed through training from ETA1–ETA4. All Au selection items showed an association with individual ETA assessments: the Au CV with ETA3 (negative association), the Au RR and Au Total sel were associated with ETA4, and the Au Int was associated with ETA2 (negative association). The weak, inverse association of the Au Int with ETA2 was noted. Of the NZ 2008 selection items, only the NZ Int was associated with ETA performance, being moderately associated with ETA3, ETA4 and Average ETA.

The 2009 Au RR was associated with all ETA performance, showing weak associations with ETA1 and ETA2, and moderate association with ETA3 and ETA4, resulting in a moderate association with Average ETA. Au Total sel was also weakly associated with ETA1 and ETA2, although not with ETA3 or ETA4. NZ RR and NZ Total sel were both strongly associated with ETA2 and ETA4; the NZ Int was strongly associated with ETA1, and was very strongly associated with ETA3. These bi-national RR and Total Sel associations were reflected in associations with the Average ETA; NZ Int was also strongly associated with Average ETA. No Au or NZ CV scores were associated with ETA performance; however the combined ANZ CV score produced a weak, negative association with ETA3.

Data for the 2010 cohort was only available for the first two ETA assessments. The Au RR was weakly, and Au Total sel was moderately, associated with ETA1. The NZ RR was moderately associated with ETA2. The Au RR and ANZ RR were weakly associated with Average ETA.

The combined years' data confirmed the relationships of the Au RR, Au Total sel and the NZ Int with ETA1, however no associations were identified between combined years' selection performance and ETA2. Noting that for ETA3 and ETA4 the combined years' data reflects the 2008 and 2009 cohorts only, relationships between combined years' selection items and ETA3 and ETA4 presented many associations: Au CV and Au RR were associated with both ETA3 and ETA4, as was NZ Int. The NZ RR was weakly associated with ETA4. The Au RR, the NZ Int and NZ Total sel were all weakly associated with Average ETA scores; the Au CV was weakly, negatively associated with Average ETA scores.

Table 20 *Correlations between GS ANZ selection and ETA performance for 2008, 2009, 2010 selection years and combined selection years; n ≥ 5*

		Pearson Correlations (*r*)														
		ETA1 %			ETA2 %			ETA3 %			ETA4 %			Average ETA %		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
2008																
	CV %	-.04	-.20	-.08	.12	-.32	.03	-.28*	-.17	-.25*	-.21	.07	-.14	-.13	-.21	-.15
	RR %	.14	-.18	.08	.11	-.13	.06	.12	.21	.14	.30*	-.00	.24*	.16	-.03	.12
	Int %	-.01	.26	.02	-.27*	.33	-.21*	.12	.43*	.15	.18	.38*	.20*	-.07	.49*	-.01
	Total sel %	.05	-.13	.02	-.09	-.15	-.11	.06	.23	.09	.24*	.23	.24*	-.01	.07	-.01
N	100	59	19	78	60	19	79	55	19	74	50	19	69	62	19	81
2009																
	CV %	-.03	-.11	-.04	-.03	.29	.02	-.19	-.21	-.21*	-.21	.24	-.16	-.16	.03	-.14
	RR %	.29**	.18	.27**	.20*	.55*	.24*	.41**	.06	.37**	.40**	.66**	.42**	.48**	.39*	.46**
	Int %	.12	.60**	.19*	.18	.37	.21*	-.03	.80*	.07	.04	.16	.07	.09	.57**	.19*
	Total sel %	.29**	.36	.30**	.26**	.65**	.33**	.16	.55	.19*	.17	.57*	.23*	.32**	.55**	.36**
N	107	77	16	93	75	15	90	68	7	75	56	14	70	79	18	97
2010																
	CV %	.10	-.04	.09	-.04	-.17	-.04	Q	Q	Q	Q	Q	Q	-.02	-.18	-.03
	RR %	.28**	-.06	.25**	.01	.43*	.09	Q	Q	Q	Q	Q	Q	.19*	.26	.22**
	Int %	.06	-.01	.01	-.00	.22	-.05	Q	Q	Q	Q	Q	Q	.04	.11	-.04
	Total sel %	.30**	-.05	.23**	-.03	.28	-.02	Q	Q	Q	Q	Q	Q	.14	.13	.09
N	140	90	16	106	110	18	128	0	0	0	0	0	0	116	18	134
Combined years																
	CV %	-.02	-.15	-.04	.01	-.05	.01	-.23**	-.09	-.23**	-.17*	.15	-.11	-.12*	-.06	-.10*
	RR %	.16**	-.05	.13*	.06	.13	.07	.27**	.07	.26**	.33**	.15	.30**	.19**	.18	.19**
	Int %	.05	.29*	.08	-.08	.21	-.06	.06	.54**	.12	.10	.28*	.12	.02	.24*	.04
	Total sel %	.13*	.07	.11*	-.01	.21	.01	.11	.28	.13	.20*	.29*	.23**	.06	.25*	.08
N	347	226	51	277	245	52	297	123	26	149	106	33	139	257	55	312

Notes. CV% = CV score as a percentage. RR% = Referee reports score as a percentage. Int % = Interview score as a percentage. Total sel % = Total selection score as a percentage. ETA % = End of Term Assessment score. Average ETA % = Mean score of ETAs 1-4.
r = Pearson Correlation. N = Number of trainees. Q = No results because n = 0 (see Table 66).
** = Correlation is significant at .01 (1-tailed). * = Correlation is significant at .05 (1-tailed). Teal = Also significant in regression analysis. Diagonal = Pearson correlation is not significant at .05 but is significant in regression analysis.

Summary of findings – Pearson correlation analyses of selection performance and assessment performance

Roughly 75 percent (n = 140) of the 186 correlations identified between selection and assessment items were weak to moderate. A similar percentage of the correlations were positive (n = 141). More correlations were identified for Au (n = 63) than for NZ (n = 51), but most correlations were identified when the data for the two countries were combined (n = 72). ANZ results, although comprised mainly of Au data, did not replicate the Au results. Correlations between Au selection items and assessments tended to be weak or very weak, while most NZ associations were moderate or strong.

Of the yearly results, the fewest significant correlations between selection and assessment performance were identified for the 2008 cohort, and most were identified for the 2009 cohort. Despite incomplete data, the 2010 cohort evinced more significant correlations than did the 2008 cohort, and showed the greatest diversity in the strength of correlations. For the 2008 cohort only, negative correlations outnumbered positive correlations.

Of the selection items, the RR, particularly in 2009 (n = 24) and in 2010 (n = 13), had most correlations with assessments (n = 56), followed by the Int (n = 47) and Total sel (n = 43). The CV had fewest correlations with assessments (n = 40), most of which were negative (n = 36). The Au RR (n = 26) showed more significant correlations than did the NZ RR (n = 7), and the NZ Int (n = 21) had more significant associations than did the Au Int (n = 11). When the two countries' data were combined, the RR maintained most correlations (n = 23), with the Total sel (n = 18), CV (n = 16) and Int (n = 15) exhibiting similar numbers of correlations to each other.

Of the assessment items, the CE presented most correlations (n = 19), and the SpecSSE, DOPS1, DOPS2 and all ETAs had 10 or more correlations with selection items. MiniCEX assessments showed fewest correlations with selection. No significant correlations were identified between examination performance and selection for the 2008 cohort; examination performance was most closely aligned to Int performance, particularly for the 2009 cohort. The GenSSE was not inversely associated with any selection item. DOPS1 and DOPS2 assessments were associated with the 2009 RR, and DOPS1 was associated with and the 2010 CV, RR and Total Sel. Three DOPS assessments were negatively associated with the 2008 Au CV and NZ CV. The RR and Int were the selection instruments most associated with ETA assessments. The 2009 selection performance was associated with all ETA assessments.

Regression analyses comparing selection with assessments

Multiple linear regression analyses were conducted to identify the extent to which scores in selection predicted scores in assessments during training. Regressions were run for the three selection items (CV, RR and Int) against each assessment (Examinations, DOPS, MiniCEX and ETAs), using a direct method; separate regressions were run for the Total sel against each assessment.

Values are reported for predictive relationships, where regression values (ANOVA) were statistically significant (i.e. $p \leq .05$). These significant predictive relationships are presented in summary tables for each assessment task. See Table 67, Table 68, Table 69 and Table 70. Results are reported where a) the 'Selection items' model is significant (and 'Total sel' $p > .05$); b) both the 'Selection items' and the 'Total sel' models are significant; c) 'Total sel' is significant and one or more selection items is significant and the 'Selection items' model $p > .05$); and d) the 'Total sel' only is significant. The model reports are differentiated in the tables by symbols as advised.

Ninety-six selection scores predicted performance in assessments. Eleven CV scores significantly predicted assessments: three CV scores were positive predictors; eight CV scores were negative predictors of assessments. See Table 71. Thirty-six RR scores were significant positive predictors of assessments; no RR scores were negative predictors of assessments. See Table 72. Sixteen Int scores were significant, positive predictors of assessments; two Int scores were significant, negative predictors of assessments. See Table 73. Thirty-one Total sel scores were significant positive predictors of assessments. No Total sel scores were negative predictors of assessments. See Table 74.

Comparing selection item predictions of assessment scores by country

The number of predictive relationships between selection scores and assessment scores varied by country. More predictive relationships were identified from Au selection than from NZ selection with most predictive relationships identified when the two countries data were combined. ETAs were most predicted by selection performance, MiniCEX were least predicted. See Figure 30.

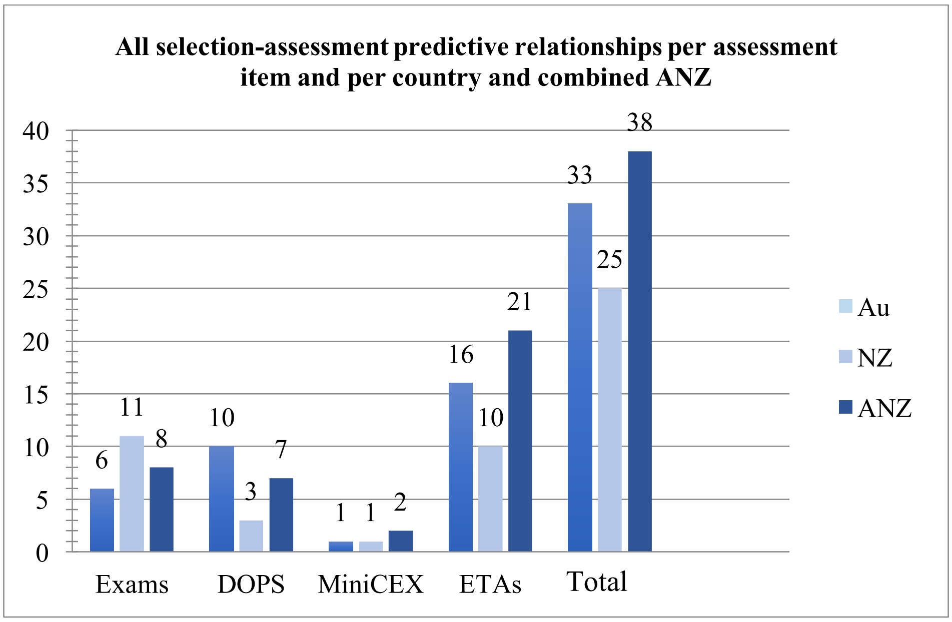

Figure 30. All selection-assessment predictive relationships per assessment item and per country and combined ANZ

The number of predictive relationships per selection item differed between Au and NZ. The CV had few predictive relationships for both countries, the RR had most predictive relationships for Au and the Int had most predictive relationships for NZ. Total sel had the second most predictive relationships for both countries. See Figure 31.

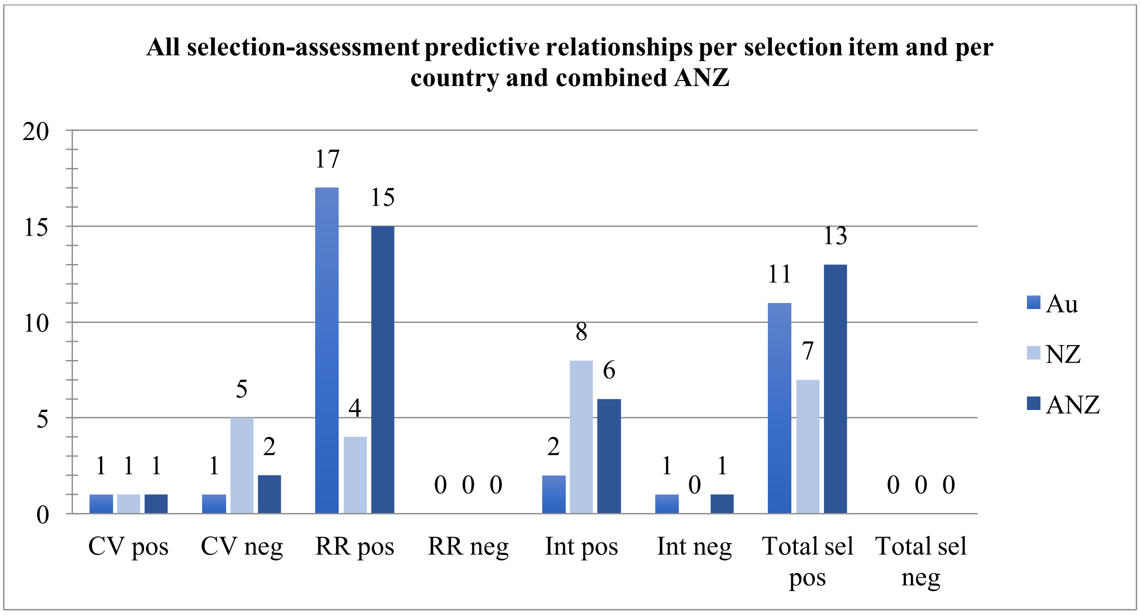

Figure 31. All selection-assessment predictive relationships per selection item and per country and combined ANZ

The number of significant correlations was greater than the number of predictions for all selection items. See Figure 32. The difference between the number of correlations and predictions was greatest for the CV and least for Total sel. The pattern of fewer predictive relationships than correlations was continued when the data was analysed per yearly cohort.

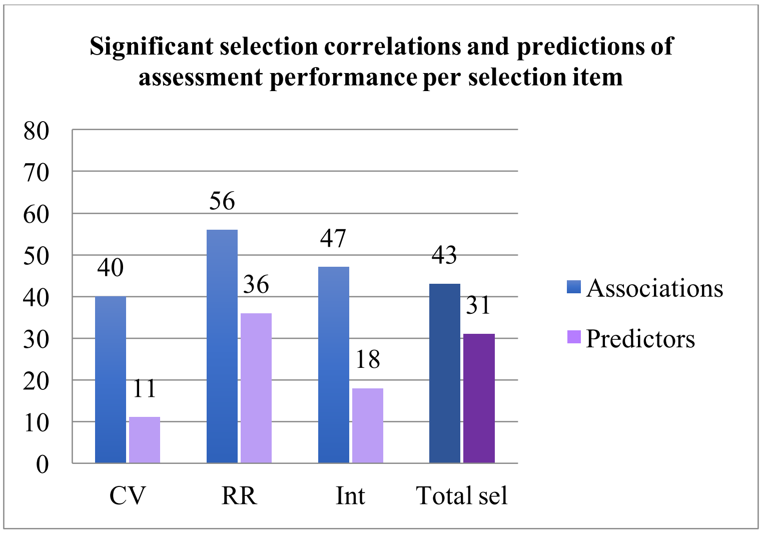

Figure 32. Significant selection correlations and predictions of assessment performance

Comparing selection item correlations with, and predictions of, assessment scores.

Fewest predictive relationships were identified for the 2008 cohort, and most were identified for the 2009 cohort, which was approximately 15% more than for the combined years' data. See Figure 33.

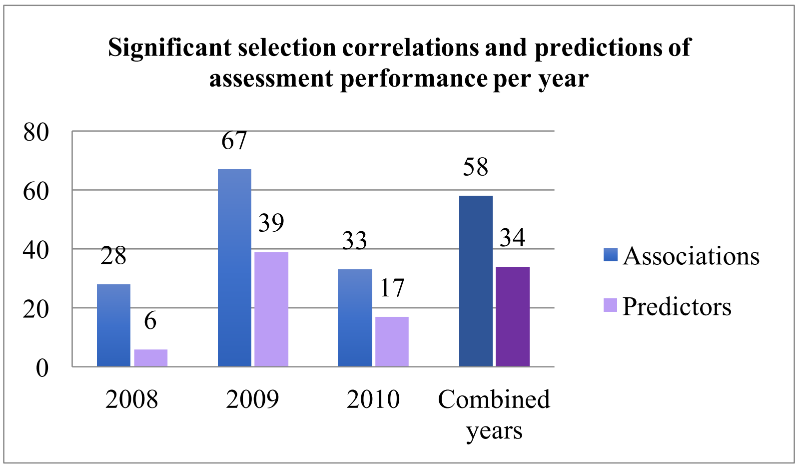

Figure 33. Significant selection correlations and predictions of assessment performance per year and combined years

Most predictive relationships were identified for the combined ANZ group (n = 38); the Au group demonstrated approximately 32% more predictive relationships between selection and assessment performance than did the NZ group. See Figure 34.

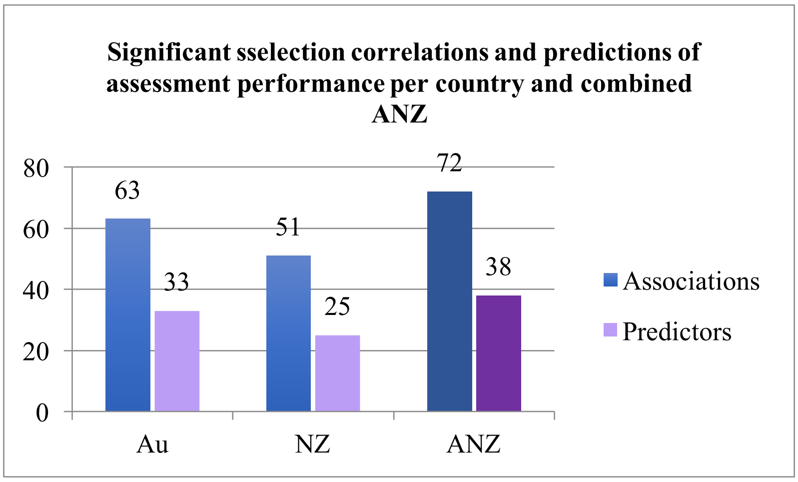

Figure 34. Significant selection correlations and predictions of assessment performance per country and combined ANZ

Regression analyses between selection scores and examination scores

Table 21, Table 22 and Table 23 present results for which significant ANOVAs were identified from linear regression models that regress each of the three selection items and Total sel on examination scores, for each of the three cohorts in each country, and for both countries combined. Au selection scores predicted performance in the CE (2009 cohort). NZ selection scores predicted performance in the GenSSE (2010 cohort), the SpecSSE (2010 cohort) and the CE (2010 cohort). Combined years' selection scores predicted performance in the GenSSE (NZ), the SpecSSE (NZ) and in the CE (Au and ANZ). See Table 67.

Regression analysis of 2008 selection scores and examination scores.

No significant predictive values of the selection scores for examinations were found for the 2008 Au or NZ cohorts.

Regression analysis of 2009 selection scores and examination scores.

Multiple linear regression analysis, used to further define relationships between variables, show that Au and ANZ RR, Int and Total sel scores significantly predicted CE performance and that the combined ANZ selection (Int) also significantly predicted SpecSSE performance. The Au Int was the strongest individual predictor, accounting for 13.5% of the variance in CE scores; the Au RR accounted for 5.4% of the variance in this examination, resulting in the Au Total sel score accounting for 7.7% of the variance in the CE.

When the Au and NZ results were combined (ANZ), the Int accounted for 13.6% of the variance in the CE scores and for 4.5% of the variance in SpecSSE scores. In addition, the ANZ RR and Total sel were predictors of performance in the CE. The model shows that ANZ selection scores accounted for 17.8% of the variance in the CE scores and 9.4% of the variance in the SpecSSE scores. See Table 21.

Table 21 *Comparing 2009 selection scores with examination scores having regression value (ANOVA) sig ≤ .05*

2009															
Examination name	**Clinical Examination**					**Specialty SSE**					**Clinical Examination**				
Country	**Au**					**ANZ**					**ANZ**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	-.028	-.253	.801	-.028	.001	-.115	-1.058	.293	-.111	.012	-.009	-.088	.930	-.009	.000
RR%	**.220**	**2.190**	**.031**	**.232**	**.054**	.119	1.166	.247	.123	.015	**.182**	**1.983**	**.050**	**.194**	**.038**
Int%	**.391**	**3.631**	**.000**	**.368**	**.135**	**.223**	**2.062**	**.042**	**.213**	**.045**	**.389**	**3.991**	**.000**	**.369**	**.136**
Total sel%	**.278**	**2.682**	**.009**	**.278**	**.077**	.123	1.179	.242	.123	.015	**.269**	**2.830**	**.006**	**.269**	**.072**
3 selection tools	$F(3,84) = 6.960$		$\boldsymbol{p < .001}$		$R^2 = .199$	$F(3,89) = 3.082$		$\boldsymbol{p = .031}$		$R^2 = .094$	$F(3,101) = 7.300$		$\boldsymbol{p < .001}$		$R^2 = .178$
Total selection	$F(3,84) = 7.194$		$\boldsymbol{p < .01}$		$R^2 = .077$	$F(1,91) = 1.390$		$p = .242$		$R^2 = .015$	$F(1,103) = 8.008$		$\boldsymbol{p = .006}$		$R^2 = .072$

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.
Bold: sig ≤ .05

Regression analysis of 2010 selection scores and examination scores.

Components of New Zealand 2010 selection performances were predictive of scores in all three examinations. Multiple linear regression analysis showed the NZ Total sel score accounted for 47.3% of the variance in the GenSSE scores and for 29.6% of the variance in the SpecSSE scores. The NZ Int accounted for 43.6% of the variance in the GenSSE and for 51.8% of the variance in the SpecSSE; and the NZ CV accounted for 33.4% of the variance in the CE scores. The models showed that the three NZ selection instruments combined to account for 54.6% of the variance in the GenSSE, for 54.0% of the variance in the SpecSSE and for 40.3% of the variance in the CE. No Au 2010 selection item predicted examination performance. See Table 22.

Table 22 *Comparing 2010 selection scores with examination scores having regression value (ANOVA) sig ≤ .05*

2010															
Examination name	**Generic SSE**					**Specialty SSE**					**Clinical Examination**				
Country	**NZ**					**NZ**					**NZ**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	.057	.305	.764	.079	.006	-.218	-1.128	.278	-.289	.084	**-.590**	**-2.740**	**.015**	**-.578**	**.334**
RR%	.304	1.700	.110	.402	.162	.115	.624	.543	.164	.027	.105	.514	.615	.131	.017
Int%	**.632**	**3.407**	**.004**	**.660**	**.436**	**.751**	**3.882**	**.002**	**.720**	**.518**	.410	1.924	.073	.445	.198
Total sel%	**.688**	**3.908**	**.001**	**.688**	**.473**	**.544**	**2.594**	**.020**	**.544**	**.296**	.037	.153	.880	. 037	.001
3 selection tools	$F(3,15) = 6.015$		$\boldsymbol{p = .007}$	$R^2 = .546$		$F(3,14) = 5.487$		$\boldsymbol{p = .011}$	$R^2 = .540$		$F(3,15) = 3.377$		$\boldsymbol{p = .046}$	$R^2 = .403$	
Total selection	$F(3,17) = 15.272$		$\boldsymbol{p = .001}$	$R^2 = .473$		$F(1,16) = 6.730$		$\boldsymbol{p = .020}$	$R^2 = .296$		$F(3,17) = .023$		$p = .880$	$R^2 = .001$	

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.
Bold: sig ≤ .05

Regression analysis of combined years' selection scores and examination scores.

Multiple regression analysis of the combined years' data confirms that the NZ Int scores, reflecting 2009 and 2010 outcomes, were the strongest predictor of examination performance, significantly predicting performance in the GenSSE and SpecSSE, respectively accounting for 17.5% and 18.7% of the variance in these examination scores. The NZ Int, the NZ CV and the NZ Total sel predicted GenSSE scores; the NZ RR, NZ Int and NZ Total sel predicted SpecSSE scores. No NZ selection item predicted performance in the CE, however Au (and ANZ) RR, Int and Total sel did predict CE performance, the Au model predicting 9.2% of the variance in CE examination scores. NZ Total sel scores accounted for 19.4% of the variance in the GenSSE and for 22.8% of the variance in the SpecSSE; Au Total sel scores accounted for 5.7% of the variance in the CE. See Table 23.

Table 23 *Comparing combined years' selection scores with examination scores having regression value (ANOVA) sig ≤ .05*

Combined years																				
Examination name	**Generic SSE**					**Specialty SSE**					**Clinical Examination**					**Clinical Examination**				
Country	NZ					NZ					Au					ANZ				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV %	**.262**	**2.109**	**.040**	**.286**	**.082**	.132	.934	.356	.152	.023	-.016	-.266	.791	-.016	.000	-.017	-.309	.757	-.017	.000
RR %	.259	1.917	.061	.262	.069	**.433**	**2.879**	**.007**	**.428**	**.183**	**.228**	**3.992**	**.000**	**.232**	**.054**	**.223**	**4.243**	**.000**	**.226**	**.051**
Int %	**.439**	**3.254**	**.002**	.418	**.175**	**.439**	**2.921**	**.006**	**.433**	**.187**	**.202**	**3.411**	**.001**	**.200**	**.040**	**.215**	**3.971**	**.000**	**.213**	**.045**
Total sel %	**.440**	**3.536**	**.001**	**.440**	**.194**	**.477**	**3.392**	**.002**	**.477**	**.228**	**.239**	**4.122**	**.000**	**.239**	**.057**	**.235**	**4.416**	**.002**	**.235**	**.055**
3 selection tools	$F(3,50) = 4.997$ **$p = .004$** $R^2 = .231$					$F(3,37) = 4.426$ **$p = .009$** $R^2 = .264$					$F(3,279) = 9.471$ **$p < .001$** $R^2 = .092$					$F(3,333) = 11.03$ **$p < .001$** $R^2 = .090$				
Total selection	$F(1,52) = 12.505$ **$p = .001$** $R^2 = .194$					$F(1,39) = 11.506$ **$p = .002$** $R^2 = .228$					$F(1,281) = 16.991$ **$p < .001$** $R^2 = .057$					$F(1,335) = 10.00$ **$p < .001$** $R^2 = .055$				

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.
Bold: sig ≤ .05

Regression analyses between selection scores and examination scores - summary

Regression analyses revealed no significant ANOVA results for the 2008 cohort with regard to examinations. Seven significant ANOVA results ($F = 3.082$ to 8.008, $p \leq .05$) were discerned for the 2009 cohort and five significant ANOVA results ($F = 3.377$ to 15.272, $p \leq .05$) identified selection scores as predictors of examination scores for the 2010 cohort. Across the years, the four selection variables (CV, RR, Int and Total sel) produced fractions of explained variances (R^2) in the range of .094 to .546. Selection scores were associated with the SpecSSE and CE scores in two cohorts (2009 and 2010) and with the GenSSE in one cohort (2010). The combined years' data showed associations between selection and all three examinations. See Table 21, Table 22, Table 23 and Table 67.

The NZ 2009 RR, Int and Total sel predicted CE performance, and the NZ 2010 selection Int and Total sel predicted performance in the GenSSE and the Spec SSE, with the NZ 2010 CV inversely predicting performance in the CE. A single significant predictive value for the Au cohorts was identified, predicting performance in the 2009 CE. See Table 21 and Table 67. Summary results are shown for regressions run with the three selection tools (CV, RR, Int) but exclude Total sel unless stated.

Regression analyses between selection scores and DOPS scores

Table 24, Table 25, Table 26 and Table 27 present results for which significant ANOVAs were identified from linear regression models that regress each of the three selection items and Total sel on DOPS scores, for each of the three cohorts in each country, and for both countries combined. Au selection scores predicted performance in in DOPS1 (2010 cohort) and DOPS3 (2010 cohort) and the Average DOPS (2009 cohort). NZ selection scores predicted performance in DOPS1 (2008 cohort) and the Average DOPS (2008 cohort). Combined years' selection scores predicted performance in DOPS1 (Au and ANZ), in DOPS2 (Au and ANZ) and in the Average DOPS (Au and NZ). See Table 68.

Regression analysis of 2008 selection scores and DOPS scores.

Multiple linear regression analysis was used to further define the relationships between 2008 selection scores and DOPS assessments. Au 2008 selection scores were not found to be significant predictors of DOPS performance. The model shows that NZ selection scores significantly predicted DOPS1 and Average DOPS performance, accounting for 59.2% of the variance in the DOPS1 scores, and for 53.8% of the variance in Average DOPS scores. The NZ CV, negatively predicted DOPS performance, accounting for 44% of the variance in the DOPS1 scores and for 48.3% of the variance in the Average DOPS scores. See Table 24.

Table 24 *Regressions between 2008 selection scores and DOPS scores having regression value (ANOVA) sig ≤ .05*

2008										
Assessment			**DOPS1**					**Average DOPS**		
Country			**NZ**					**NZ**		
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	**-.579**	**-2.797**	**.019**	**-.663**	**.440**	**-.674**	**-3.057**	**.012**	**-.695**	**.483**
RR%	.041	0.174	.865	.055	.003	.057	0.230	.823	.072	.005
Int%	.481	2.072	.065	.548	.300	.248	1.002	.340	.302	.091
Total sel%	-.313	-1.143	.275	-.313	.098	-.389	-1.463	.169	-.389	.151
3 selection tools	$F(3,10) = 4.845$	$\boldsymbol{p = .025}$	$R^2 = .592$			$F(3,10) = 3.887$	$\boldsymbol{p = .044}$	$R^2 = .538$		
Total selection	$F(1,12) = 1.307$	$p = .275$	$R^2 = .098$			$F(1,12) = 2.141$	$p = .169$	$R^2 = .151$		

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.
Bold: sig ≤ .05.

Regression analysis of 2009 selection scores and DOPS scores.

Multiple regression analyses show that the Au 2009 RR scores predicted 7.4% of the variance in the Average DOPS score; the Au 2009 Total sel scores predicted 7.2% of the variance in Average DOPS score. The ANZ Total sel predicted 5.4% of the variance in the

Average DOPS scores. Neither Au nor NZ 2009 selection scores predicted performance in any individual DOPS assessment. See Table 25.

Table 25 *Regressions between 2009 selection scores and DOPS scores having regression value (ANOVA) sig ≤ .05*

2009										
Assessment	**Average DOPS**					**Average DOPS**				
Country	**Au**					**ANZ**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	.113	.825	.412	.104	.011	.088	.712	.479	.084	.007
RR%	**.272**	**2.226**	**.030**	**.272**	**.074**	.206	1.805	.075	.208	.043
Int%	.136	.996	.323	.125	.016	.156	1.263	.211	.147	.022
Total sel%	**.268**	**2.222**	**.030**	**.268**	**.072**	**.232**	**2.054**	**.043**	**.232**	**.054**
3 selection tools	$F(1,66) = 1.911$ $p = .137$ $R^2 = .085$					$F(3,72) = 1.555$ $p = .208$ $R^2 = .061$				
Total selection	$F(1,64) = 4.937$ $\boldsymbol{p = .030}$ $R^2 = .072$					$F(1,74) = 4.219$ $\boldsymbol{p = .043}$ $R^2 = .054$				
Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%. **Bold**: sig ≤ .05.										

Regression analysis of 2010 selection scores and DOPS scores.

Multiple linear regression analysis link Au (and ANZ) 2010 selection scores with DOPS1 performance and with DOPS3 performance. Au selection accounted for 9.7% of the variance in DOPS1 scores, and for 24.1% of the variance in DOPS3 scores. Within the model, Au CV, RR and Total sel all predicted DOPS1 performance, with Au Total sel being the strongest individual predictor, accounting for 8.2% of the variance in this assessment. The Au 2010 Int accounted for 19.8% of the variance in DOPS3, the relationship being negative, indicating that high scores in the Au 2010 Int were likely to be associated with low scores in DOPS3. The 2010 ANZ result for DOPS1 was consistent with the Au result; no NZ scores were available for DOPS3, so the ANZ relationship only reflected Au performance in this assessment. See Table 26.

Table 26 *Regressions between 2010 Selection scores and DOPS scores having regression value (ANOVA) sig ≤ .05*

2010										
Assessment	**DOPS1**					**DOPS3**				
Country	**Au**					**Au**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	**.267**	**2.698**	**.008**	**.263**	**.069**	.008	.049	.961	.009	.000
RR%	**.216**	**2.182**	**.032**	**.215**	**.046**	.109	.648	.522	.117	.014
Int%	.095	.949	.345	.095	.009	**-.450**	**-2.720**	**.011**	**-.445**	**.198**
Total sel%	**.287**	**2.999**	**.003**	**.287**	**.082**	-.070	-.397	.694	-.070	.005
3 selection tools	$F(3,98) = 3.525$		***p* = .018**		$R^2 = .097$	$F(3,30) = 3.179$		***p* = .038**		$R^2 = .241$
Total selection	$F(1,100) = 8.993$		***p* = .003**		$R^2 = .083$	$F(1,32) = 0.158$		$p = .694$		$R^2 = .005$
Assessment	**DOPS1**					**DOPS3**				
Country	**ANZ**					**ANZ**[a]				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	**.245**	**2.695**	**.008**	**.244**	**.060**	.008	.049	.961	.009	.000
RR%	**.194**	**2.104**	**.038**	**.193**	**.037**	.109	.648	.522	.117	.014
Int%	.089	.959	.340	.089	.008	**-.450**	**-2.720**	**.011**	**-.445**	**.198**
Total sel%	**.254**	**2.842**	**.005**	**.254**	**.065**	-.070	-.397	.694	-.070	.005
3 selection tools	$F(3,115) = 3.475$		***p* = .018**		$R^2 = .083$	$F(3,30) = 3.179$		***p* = .038**		$R^2 = .241$
Total selection	$F(1,117) = 8.078$		***p* = .005**		$R^2 = .065$	$F(1,32) = 0.158$		$p = .694$		$R^2 = .005$

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.

Bold: sig ≤ .05.

[a] 2010 ANZ results for DOPS3 = Au results for DOPS3 as no NZ scores available.

Regression analysis of combined years' selection scores and DOPS scores.

Regression analysis results for the combined cohorts confirmed that Au (and ANZ) selection scores predicted DOPS1 and DOPS2 performance and that Au selection predicted Average DOPS performance; the NZ CV predicted Average DOPS scores. Within the model, Au RR and Au Total sel predicted 2.1% and 3.5% respectively of DOPS1 performance. The Au Total sel result was tempered in the ANZ Total sel, which predicted 1.7% of DOPS1 performance; ANZ CV scores predicted 2.7% of DOPS2 performance. Although only directly predicting 2.1% of DOPS1 performance, the strength of the Au RR predictive relationship increased to 3.5% in the Average DOPS, indicating that the Au RR was consistently associated with DOPS performance. See Table 27.

Table 27 *Regressions between combined years selection scores and DOPS scores having regression value (ANOVA) sig ≤ .05*

Combined years															
Assessment			**DOPS1**					**DOPS2**					**Average DOPS**		
Country			**Au**					**Au**					**Au**		
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	.090	1.275	.204	.088	.008	-.133	-1.740	.084	-.131	.017	-.003	-.044	.965	-.003	.000
RR%	**.145**	**2.121**	**.035**	**.146**	**.021**	.124	1.670	.097	.126	.016	**.188**	**2.774**	**.006**	**.188**	**.035**
Int%	.119	1.691	.092	.117	.014	.103	1.344	.181	.102	.010	.051	.729	.467	.050	.003
Total sel%	**.186**	**2.740**	**.007**	**.186**	**.035**	.055	.735	.464	.055	.003	**.138**	**2.022**	**.044**	**.138**	**.019**
3 selection tools	F(3,207) = 2.615		**p = .052**	R^2 = .037		F(3,173) = 2.945		**p = .034**	R^2 = .049		F(3,210) =2.689		**p = .047**	R^2 = .037	
Total selection	F(1,209) = 7.509		**p = .007**	R^2 = .035		F(1,175) = .540		**p = .464**	R^2 = .003		F(1,212) =.4.088		**p = .044**	R^2 = .019	
Assessment			**Average DOPS**					**DOPS1**					**DOPS2**		
Country			**NZ**					**ANZ**					**ANZ**		
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	**-.352**	**-2.375**	**.023**	**-.364**	**.132**	.055	.847	.398	.054	.003	**-.167**	**-2.536**	**.019**	**-.165**	**.027**
RR%	-.178	-1.129	.266	-.182	.033	.087	1.375	.170	.087	.008	.101	1.452	.148	.102	.010
Int%	.140	.890	.379	.145	.021	.114	1.757	.080	.111	.012	.077	1.082	.281	.076	.006
Total sel%	-.217	-1.390	.172	-.217	.047	**.132**	**2.098**	**.037**	**.132**	**.017**	.019	.274	.785	.019	.000
3 selection tools	F(3,37) = 2.859		**p = .050**	R^2 = .188		F(3,248) = 1.586		p = .193	R^2 = .019		F(3,200) = 3.455		**p = .017**	R^2 = .049	
Total selection	F(1,39) = 0.1.931		p = .172	R^2 = .047		F(1,250) = 4.400		**p = .037**	R^2 = .017		F(1,202) = .075		p = .785	R^2 = .000	

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.

Bold: sig ≤ .05.

Regression analyses between selection scores and DOPS scores - summary

Regression analyses revealed two Au and one NZ significant ANOVA results ($F = 3.179$ to 4.845, $p < .05$) for yearly selection scores as predictors of individual DOPS scores (Au 2010 cohort predicted DOPS1 and DOPS3; NZ 2008 cohort predicted DOPS1). Au RR and Total sel (2009 cohort) and NZ CV (2008 cohort, inverse relationship) were identified as predictors of Average DOPS scores. The four selection tool variables produced fractions of explained variances (R^2) in the range of .030 to .592. The 2009 cohort selection scores did not predict any individual DOPS assessment score. See Table 68.

The combined years' Au and ANZ selection scores predicted two DOPS assessments (DOPS1 and DOPS2), and also predicted the average DOPS scores. Results are shown for regressions conducted with all three selection tools excluding total selection unless otherwise stated (e.g. see 2009 Au – Average DOPS). A summary is presented in Table 68.

Regression analysis between selection scores and MiniCEX scores

Table 28 and Table 29 present results for which significant ANOVAs were identified from linear regression models that regress each of the three selection items and Total sel on MiniCEX scores, for each of the three cohorts in each country, and for both countries combined. Au selection scores predicted performance in the Average MiniCEX (2009 cohort). NZ selection scores did not predict performance in any individual MiniCEX. Combined years' NZ selection scores predicted scores in the Average MiniCEX. See Table 69.

Regression analysis of 2008 selection scores and MiniCEX scores.

No significant predictive values of the selection scores for MiniCEX assessments were found for the 2008 Au or NZ cohorts.

Regression analysis of 2009 selection scores and MiniCEX scores.

Pearson correlations indicated associations between selection and individual MiniCEX assessments, however, multiple regression analysis revealed that the 2009 cohort's selection scores did not predict any individual MiniCEX assessment; the Au 2009 selection scores did predict the Average MiniCEX scores. The Au RR accounted for 5.7% of the variance in the Average MiniCEX scores. This was confirmed and slightly strengthened in the ANZ scores, with the RR predicting 6.2% of the variance in Average MiniCEX. See Table 28.

Table 28 *Regressions between 2009 selection scores and MiniCEX scores having regression value (ANOVA) sig ≤ .05*

2009										
Assessment	**Average MiniCEX**					**Average MiniCEX**				
Country	**Au**					**ANZ**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	-.220	-1.672	.099	-.205	.042	-.193	-1.636	.106	-.187	.035
RR%	**.233**	**1.972**	**.053**	**.239**	**.057**	**.241**	**2.198**	**.031**	**.248**	**.062**
Int%	.049	.377	.708	.047	.002	.079	.674	.502	.078	.006
Total sel%	.098	.799	.427	.098	.010	.111	.973	.334	.111	.012
3 selection tools	$F(3,64) = 3.052$	$\boldsymbol{p = .035}$	$R^2 = .125$			$F(3,74) = 3.279$	$\boldsymbol{p = .026}$	$R^2 = .117$		
Total selection	$F(1,66) = 0.638$	$p = .427$	$R^2 = .010$			$F(1,76) = .947$	$p = .334$	$R^2 = .012$		

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.
Bold: sig ≤ .05.

Regression analysis of 2010 selection scores and MiniCEX scores.

No significant predictive values of the selection scores for MiniCEX assessments were found for the 2010 Au or NZ cohorts.

Regression analysis of combined years' selection scores and MiniCEX scores.

Multiple regression analysis of the combined years' data identified that the NZ CV accounted for 11.9% of the variance in Average MiniCEX scores and that ANZ RR accounted for 1.8% of the variance in Average MiniCEX scores. See Table 29.

Table 29 *Regressions between combined years' selection scores and MiniCEX scores having regression value (ANOVA) sig ≤ .05*

Combined years										
Assessment	**Average MiniCEX**					**Average MiniCEX**				
Country	**NZ**					**ANZ**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	**-.323**	**-2.233**	**.032**	**-.345**	**.119**	-.114	-1.797	.073	-.112	.013
RR%	.218	1.406	.168	.225	.051	**.134**	**2.149**	**.033**	**.134**	**.018**
Int%	.371	2.388	.022	.365	.133	.045	.704	.482	.044	.002
Total sel%	.072	.452	.654	.072	.005	.046	.736	.463	.046	.002
3 selection tools	$F(3,37) = 3.778$	$\boldsymbol{p = .018}$	$R^2 = .235$			$F(3,253) = 2.899$	$\boldsymbol{p = .036}$	$R^2 = .033$		
Total selection	$F(1,39) = .204$	$p = .654$	$R^2 = .072$			$F(1,255) = .541$	$p = .463$	$R^2 = .002$		

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.
Bold: sig ≤ .05.

Regression analyses between selection scores and MiniCEX scores - summary

Discounting $n < 5$, regression analyses revealed no significant ANOVA results for any single cohort's selection scores as predictors of individual MiniCEX scores, however the 2009 Au and ANZ selection indicated prediction for average MiniCEX scores. The combined years' data also indicated associations between selection and average MiniCEX scores. A summary is presented in Table 69.

Regression analyses between selection scores and ETA scores

Table 30, Table 31, Table 32, Table 33, Table 34, Table 35 and Table 36 present results for which significant ANOVAs were identified from linear regression models that regress each of the three selection items and Total sel on ETA scores, for each of the three cohorts in each country, and for both countries combined. Au selection scores predicted performance in ETA1

(2009 cohort and 2010 cohort), ETA2 (2009 cohort), ETA3 (2009 cohort, ETA4 (2008 cohort and 2009 cohort), and the Average ETA (2009 cohort). NZ selection scores predicted performance in ETA2 (2009 cohort), ETA3 (2008 cohort), ETA4 (2009 cohort), and the Average ETA (2009 cohort). Combined years' selectin scores predicted performance in ETA3 (Au and NZ), in ETA4 (Au) and in the Average ETA (Au and NZ). See Table 70.

Table 30 *Regressions between 2008 selection scores and ETA scores having regression value (ANOVA) sig ≤ .05*

2008															
Assessment			**ETA3**					**ETA4**					**ETA4**		
Country			**NZ**					**Au**					**ANZ**		
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	-.259	-1.280	.220	-.314	.099	-.179	-1.265	.212	-.183	.033	-.099	-.830	.409	-.102	.010
RR%	.443	2.058	.057	.469	.220	**.320**	**2.367**	**.022**	**.329**	**.108**	**.266**	**2.268**	**.027**	**.271**	**.073**
Int%	**.625**	**2.876**	**.012**	**.596**	**.355**	.151	1.064	.293	.155	.024	.209	1.742	.086	.211	.045
Total sel%	.230	.974	.344	.230	.053	.240	1.710	.094	.240	.058	**.237**	**2.000**	**.050**	**.237**	**.056**
3 selection tools	$F(3,15) = 3.360$		**p = .047**		$R^2 = .402$	$F(3,46) = 2.967$		**p = .042**		$R^2 = .162$	$F(3,65) = 2.925$		**p = .040**		$R^2 = .119$
Total selection	$F(1,17) = 0.949$		**p = .344**		$R^2 = .053$	$F(1,48) = 2.923$		p = .094		$R^2 = .057$	$F(1,67) = 3.999$		**p = .050**		$R^2 = .056$

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.

Bold: sig ≤ .05.

Regression analysis of 2008 selection scores and ETA scores.

No 2008 CV scores predicted scores in ETAs. Multiple linear regression analysis identified the NZ Int as a predictor of ETA3, predicting 35.5% of the variance in this assessment, and the Au RR as predicting 10.8% of the variance in ETA4. When Au and NZ scores were combined, ANZ Int scores predicted 7.3% of the variation in ETA4 and ANZ Total sel predicted 5.6% of the variation in ETA4 performance. There was little synchronicity between Au and NZ regarding the significance of selection items predicting ETA performance. See Table 30.

Regression analysis of 2009 selection scores and ETA scores.

No 2009 CV scores predicted scores in ETAs. Regression results for the 2009 cohort confirmed that Au RR predicted ETA1, ETA2, ETA3, ETA4 and Average ETA scores, and the NZ RR and NZ Total sel predicted performance in the ETA2 and ETA4. The Au RR predicted 9.4% of the variance in ETA1, 5.2% of the variance in ETA2, 15.8% of ETA3 and 14.9% of ETA4. The Au Total sel predicted 8.1% of the variance in ETA1 performance and 6.9% of the variance in ETA2. See Table 31.

Table 31 *Regressions between 2009 Au and NZ selection scores with ETA1-ETA4 scores having regression value (ANOVA) sig ≤ .05*

2009																				
	ETA1					**ETA2**					**ETA3**					**ETA4**				
	Au					**Au**					**Au**					**Au**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV %	.074	.595	.554	.069	.005	.098	.770	.444	.091	.008	-.198	-1.612	.112	-.198	.039	-.194	-1.343	.185	-.183	.033
RR %	**.308**	**2.754**	**.007**	**.307**	**.094**	**.225**	**1.960**	**.054**	**.227**	**.052**	**.389**	**3.456**	**.001**	**.397**	**.158**	**.379**	**3.015**	**.004**	**.386**	**.149**
Int %	.174	1.399	.166	.162	.026	.243	1.915	.060	.222	.049	-.090	-.731	.468	-.091	.008	-.065	-.454	.652	-.063	.004
Total sel %	**.285**	**2.579**	**.012**	**.285**	**.081**	**.263**	**2.325**	**.023**	**.263**	**.069**	.159	1.305	.197	.159	.025	.172	1.285	.204	.172	.030
3 selection tools	*F*(3,73) = 2.937		***p* = .039**		R^2 = .108	*F*(3,71) = 2.224		*p* = .093		R^2 = .086	*F*(3,64) = 5.325		***p* = .002**		R^2 = .200	*F*(3,52) = 4.092		***p* = .011**		R^2 = .191
Total selection	*F*(1,75) = 6.651		***p* = .012**		R^2 = .081	*F*(1,73) = 5.405		***p* = .023**		R^2 = .069	*F*(1,66) = 1.702		*p* = .197		R^2 = .025	*F*(1,54) = 1.652		*p* = .204		R^2 = .172

Assessment	**ETA2**					**ETA4**				
Country	**NZ**					**NZ**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	.190	.863	.407	.252	.064	.228	1.004	.339	.303	.092
RR%	**.528**	**2.399**	**.035**	**.586**	**.343**	**.646**	**2.830**	**.018**	**.667**	**.445**
Int%	.371	1.704	.116	.457	.209	.065	.284	.782	.089	.008
Total sel%	**.654**	**3.114**	**.008**	**.654**	**.428**	**.573**	**2.424**	**.032**	**.573**	**.328**
3 selection tools	*F*(3,11) = 3.38		*p* = .057		R^2 = .481	*F*(3,10) = 3.167		*p* = .073		R^2 = .487
Total selection	*F*(1,13) = 9.699		***p* = .008**		R^2 = .427	*F*(1,12) = 5.875		***p* = .032**		R^2 = .329

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.

Bold: sig ≤ .05.

When Au and NZ scores were combined, the RR predicted all individual ETA assessments and the Int was revealed as predicting 4.8% of variance in ETA1 and 7.2% of variation in ETA2. Au RR, Au Total sel and ANZ RR and Total sel predicted Average ETA performance; the NZ Int predicted the Average NZ ETA performance, although the NZ Int was not identified as a significant predictor of individual ETA scores. These outcomes led the ANZ RR, Int and Total sel to predict the Average ETA scores, with the ANZ RR being the strongest predictor of the three items, predicting 22.8% of the variance in Average ETA scores. See Table 32, and Table 33.

Table 32 *Regressions between 2009 ANZ selection scores with ETA1-ETA4 scores having regression value (ANOVA) sig ≤ .05*

2009	**ETA1 ANZ**					**ETA2 ANZ**					**ETA3 ANZ**					**ETA4 ANZ**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV %	.068	.621	.536	.066	.004	.140	1.276	.206	.136	.018	-.171	-1.418	.161	-.166	.028	-.123	-.999	.322	-.122	.015
RR %	**.287**	**2.864**	**.005**	**.291**	**.085**	**.271**	**2.671**	**.009**	**.277**	**.077**	**.357**	**3.270**	**.002**	**.362**	**.131**	**.417**	**3.770**	**.000**	**.421**	**.177**
Int %	**.231**	**2.128**	**.036**	**.220**	**.048**	**.283**	**2.587**	**.011**	**.269**	**.072**	.021	.174	.862	.021	.000	.003	.025	.980	.003	.000
Total sel %	**.301**	**3.006**	**.003**	**.301**	**.091**	**.334**	**3.319**	**.001**	**.334**	**.112**	.191	1.662	.101	.191	.036	**.233**	**1.976**	**.050**	**.233**	**.054**
3 selection tools	$F(3,89) = 3.996$ **p = .010** $R^2 = .119$					$F(3,86) = 4.181$ **p = .008** $R^2 = .127$					$F(3,71) = 4.862$ **p = .004** $R^2 = .170$					$F(3,66) = 5.458$ **p = .002** $R^2 = .199$				
Total selection	$F(1,91) = 9.035$ **p = .003** $R^2 = .090$					$F(1,88) = 11.017$ **p = .001** $R^2 = .111$					$F(1,73) = 2.761$ $p = .101$ $R^2 = .036$					$F(1,68) = 3.905$ **p = .050** $R^2 = .054$				

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.

Bold: sig ≤ .05.

Table 33 *Regressions between 2009 Au, NZ, and ANZ selection scores with Average ETA scores having regression value (ANOVA) sig ≤ .05*

2009															
Assessment	**Average ETA**					**Average ETA**					**Average ETA**				
Country	**Au**					**NZ**					**ANZ**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	-.074	-.657	.513	-.076	.006	-.006	-.028	.978	-.007	.000	-.024	-.245	.807	-.025	.001
RR%	**.483**	**4.782**	**.000**	**.483**	**.233**	.362	1.819	.090	.437	.191	**.470**	**5.234**	**.000**	**.477**	**.228**
Int%	.099	.879	.382	.101	.010	**.547**	**2.759**	**.015**	**.593**	**.352**	**.210**	**2.150**	**.034**	**.218**	**.048**
Total sel%	**.319**	**2.955**	**.004**	**.319**	**.102**	**.546**	**2.604**	**.019**	**.546**	.298	**.369**	**3.872**	**.000**	**.369**	**.136**
3 selection tools	$F(3,75) = 8.532$		$\boldsymbol{p = .004}$		$R^2 = .254$	$F(3,14) = 3.846$		$\boldsymbol{p = .034}$		$R^2 = .452$	$F(3,93) = 10.818$		$\boldsymbol{p = .000}$		$R^2 = .259$
Total selection	$F(1,77) = 8.731$		$\boldsymbol{p = .004}$		$R^2 = .102$	$F(1,16) = 6.783$		$\boldsymbol{p = .019}$		$R^2 = .298$	$F(1,95) = 14.992$		$\boldsymbol{p = .000}$		$R^2 = .136$

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.

Bold: sig ≤ .05.

Regression analysis of 2010 selection scores and ETA scores.

The 2010 cohort's Au RR predicted 10.2% of the variance in ETA1 scores and Au Total sel predicted 9.1% of the variance in ETA1 scores. These associations carried through to the ANZ RR, which predicted 9.1% of the variance in ETA1 scores and ANZ Total sel predicting 5.1% of the variance in ETA1 scores. See Table 34.

Table 34 *Regressions between 2010 selection scores with ETA scores having regression value (ANOVA) sig ≤ .05*

2010										
Assessment	**ETA 1**					**ETA 1**				
Country	**Au**					**ANZ**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	.165	1.577	.118	.168	.028	.130	1.344	.182	.132	.017
RR%	**.329**	**3.133**	**.002**	**.320**	**.102**	**.282**	**2.870**	**.005**	**.273**	**.075**
Int%	.160	1.513	.134	.161	.026	.087	.877	.383	.087	.008
Total sel%	**.302**	**2.977**	**.004**	**.302**	**.091**	**.226**	**2.361**	**.020**	**.226**	**.051**
3 selection tools	$F(3,86) = 3.789$	$\boldsymbol{p = .013}$	$R^2 = .117$			$F(3,102) = 3.036$	$\boldsymbol{p = .033}$	$R^2 = .082$		
Total selection	$F(1,88) = 8.862$	$\boldsymbol{p = .004}$	$R^2 = .091$			$F(1,104) = 5.572$	$\boldsymbol{p = .020}$	$R^2 = .051$		

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.
Bold: sig ≤ .05.

Regression analysis of combined years' selection scores and ETA score.

Regression results for the combined cohorts showed no predictive relationships between selection and ETA1 or ETA2 performance. Two Au selection instruments predicted performance in ETA3 (Au CV 4.5% of variance (inverse relationship); Au RR, 7.0% of variance); the NZ Int predicted 34.8% of variance in ETA3. The Au RR predicted 10.5% of variance in ETA4 and 3.4% of the variance in the Average ETA. The Au Total sel predicted 4.0% of the variance in ETA4.

The NZ RR was not shown to predict any individual ETA assessment outcomes, but did predict 9.2% of the variance in the Average ETA scores. The NZ Int predicted 11.8% of the variance in the Average ETA scores. Combined ANZ scores showed the CV as a negative predictor of ETA3 and confirmed the RR as a predictor of ETA3, ETA4 and Average ETA. ANZ Total sel predicted 5.0% of the variance in ETA4. See Table 35 and Table 36.

Table 35 *Regressions between combined years' Au and NZ selection scores with ETA scores having regression value (ANOVA) sig ≤ .05*

Combined years															
Assessment	**ETA3**					**ETA4**					**Average ETA**				
Country	**Au**					**Au**					**Au**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	**-.218**	**-2.354**	**.020**	**-.211**	**.045**	-.119	-1.178	.242	-.116	.013	-.118	-1.840	.067	-.115	.013
RR%	**.259**	**3.002**	**.003**	**.265**	**.070**	**.321**	**3.458**	**.001**	**.324**	**.105**	**.183**	**2.984**	**.003**	**.184**	**.034**
Int%	-.020	-.211	.833	-.019	.000	.056	.561	.576	.055	.003	-.010	-.153	.879	-.010	.000
Total sel%	.105	1.162	.247	.105	.011	**.200**	**2.080**	**.040**	**.200**	**.040**	.059	.948	.344	.059	.003
3 selection tools	$F(3,119) = 5.404$ **$p = .002$** $R^2 = .120$					$F(3,102) = 5.178$ **$p = .002$** $R^2 = .132$					$F(3,253) = 4.223$ **$p = .006$** $R^2 = .048$				
Total selection	$F(1,121) = 1.351$ $p = .247$ $R^2 = .011$					$F(1,104) = 4.328$ **$p = .040$** $R^2 = .040$					$F(1,255) = .899$ $p = .344$ $R^2 = .004$				

Assessment	**ETA3**					**Average ETA**				
Country	**NZ**					**NZ**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	-.076	-.442	.663	-.094	.009	-.079	-.612	.543	-.085	.007
RR%	.258	1.426	.168	.291	.085	**.318**	**2.269**	**.028**	**.303**	**.092**
Int%	**.618**	**3.431**	**.002**	**.590**	**.348**	**.365**	**2.611**	**.012**	**.343**	**.118**
Total sel%	.275	1.400	.174	.275	.076	.252	1.899	.063	.252	.064
3 selection tools	$F(3,22) = 4.064$ **$p = .019$** $R^2 = .357$					$F(3,51) = 2.979$ **$p = .040$** $R^2 = .149$				
Total selection	$F(1,24) = 1.959$ $p = .174$ $R^2 = .075$					$F(1,53) = 3.606$ $p = .063$ $R^2 = .064$				

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.
Bold: sig ≤ .05.

Table 36 *Regressions between combined years' ANZ selection scores with ETA scores having regression value (ANOVA) sig ≤ .05*

Combined years															
Assessment	**ETA3**					**ETA4**					**Average ETA**				
Country	**ANZ**					**ANZ**					**ANZ**				
	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2	ß	t	Sig.	*pr*	spr^2
CV%	**-.186**	**-2.227**	**.027**	**-.182**	**.033**	-.049	-.565	.573	-.049	.002	-.096	-1.671	.096	-.095	.009
RR%	**.242**	**3.076**	**.003**	**.247**	**.061**	**.298**	**3.641**	**.000**	**.299**	**.089**	**.183**	**3.277**	**.001**	**.184**	**.034**
Int%	.061	.739	.461	.061	.004	.117	1.353	.178	.116	.013	.026	.443	.658	.025	.001
Total sel%	.132	1.609	.110	.132	.017	**.224**	**2.690**	**.008**	**.224**	**.050**	.080	1.407	.160	.080	.006
3 selection tools	$F(3,145) = 6.045$ **$p = .001$** $R^2 = .111$					$F(3,135) = 5.482$ **$p = .001$** $R^2 = .109$					$F(3,308) = 4.754$ **$p = .003$** $R^2 = .044$				
Total selection	$F(1,147) = 2.588$ $p = .110$ $R^2 = .017$					$F(1,137) = 7.236$ **$p = .008$** $R^2 = .050$					$F(1,310) = 1.980$ $p = .160$ $R^2 = .006$				

Note. Table presents regression conducted for 3 selection variables (CV, RR, Int) and a separate regression conducted for Total sel%.
Bold: sig ≤ .05.

Regression analysis between selection scores and ETA scores - summary

Regression analyses revealed nine significant ANOVA results (F = 2.937 to 9.699, $p < .05$) for yearly selection scores as predictors of individual ETA scores (2008: Au RR predicted ETA4; NZ Int predicted ETA3. 2009: Au RR predicted ETA1, ETA2, ETA3, ETA4; NZ RR and NZ Total sel predicted in ETA2 and ETA4. 2010: Au RR and Au Total sel predicted ETA1). Au (RR and Total Sel) and NZ (Int and Total sel) scores predicted the average ETA in 2009. The four selection tool variables (CV, RR, Int and Total sel) produced fractions of explained variances (R^2) in the range of .069 to .452.

Neither the Au CV nor the NZ CV predicted ETA performance for any cohort, however, when the yearly cohorts' data were combined, the Au CV was inversely correlated to the ETA3, predicting 4.5% of the variance in this assessment. The combined years' data showed selection predicted ETA3 (Au CV (inverse relationship), Au RR and NZ Int), ETA4 (Au RR and Au Total sel) and Average ETA (Au RR, NZ RR and NZ Int). A summary is presented in Table 70.

Summary to the Results

Intake of GS trainees was approximately 100 for 2008 and 2009, rising to 140 in 2010; approximately 80% trained in Au, 20% in NZ. Mean CV scores were relatively low (43%-54%) and RR and Int mean scores were high (RR 76%-88%; Int 73%-83%) contributing to high mean Total sel scores (73%-77%). Mean RR scores and Total sel scores increased each year. Mean assessment scores were high for written examinations (72%-76%), lower for the CE (64%-70%) and high for work-based assessments (79%-92%). Selection and assessment scores tended to have narrow range of scores. Scores tended to be consistent between Au and NZ and across the three yearly cohorts.

Eight significant correlations were identified for performance across the three selection instruments. Performance in the Au Int was negatively correlated with the Au CV (2009, 2010) and the Au RR (2010). Thirty-two significant correlations were identified between performance in examinations; eight significant correlations were identified within DOPS and nine within the

MiniCEX assessments and 21 significant correlations were identified between performance in ETAs.

Pearson correlation analyses identified 186 significant associations between performance in selection and in subsequent assessments. Most significant correlations were weak or moderate positive associations. Most RR and Int correlations with assessment items were positive, however almost all CV significant correlations were negative. Most correlations were identified between the RR and subsequent performance (n = 56) with fewest between the CV and subsequent performance (n = 40). Total sel demonstrated 43 significant correlations with subsequent performance, and the Int demonstrated 47. Most of the significant correlations were with the CE, the first two DOPS and the ETAs. Correlations tended to be weaker for Au than for NZ. More significant correlations were identified between selection in 2009 than for 2008 or 2010.

Regression analyses identified 96 instances where performance in selection significantly predicted performance in assessments. Most predictive relationships were identified between the RR and subsequent performance (n = 36) with fewest between the CV and subsequent performance (n = 11). Analysis revealed 31 significant predictive relationships for Total sel performance and assessment performance, and 18 for Int performance. Most CV predictions of assessment performance were negative (n = 3 pos; n = 8 neg), all RR and Total sel predictions of assessment performance were positive, and most Int predictions were positive (n = 16 pos; n = 2 neg). See Figure 31. Most of the predictive relationships were for performance in the ETAs (n = 47) and the examinations (n = 25). Fewest predictive relationships were identified for the MiniCEX assessments (n = 4). See Figure 30. More predictive relationships were identified for selection for the 2009 cohort (n = 39) than for 2008 (n = 6) or 2010 (n = 17). The combined years' data disclosed more predictive relationships (n = 34) than for either 2008 or 2010. See Figure 33. More predictive relationships were identified for Au (n = 33) than for NZ (n = 25), with most predictive relationships identified for combined ANZ performance (n = 38). See Figure 34.

CHAPTER 5: DISCUSSION

This study aimed to highlight patterns of candidate and trainee performance and relationships between selection scores and assessment scores that were revealed in the analysis. The Discussion reviews patterns of performance and the extent to which selection scores predict early assessment scores. The Discussion also reviews the context for Australian and New Zealand (ANZ) surgical selection, training and assessment, and appraises key individual roles in RACS' surgical selection, training and assessment. Themes presented in the Literature Review are reconsidered and are linked and applied to recommendations. The Discussion is presented in four parts: Part one presents a discussion of the findings, as patterns of performance and relationships; Parts two, three, and four elaborate on themes identified in the Introduction and Literature Review, in the context of RACS' surgical education and training (SET) program. In addition, the Discussion presents implications for practice, limitations of the study and recommendations for future studies.

Part 1 Reflection on the strategic importance of underpinning selection systems with theory

This section presents a discussion of the findings. Firstly, I consider the implications of scores in selection activities and the interrelationships between selection items. Secondly, I review performance in assessment items and interrelationships between assessment items. Thirdly, I comment on relationships identified between trainee performance in selection and in subsequent assessments. Fourthly, I review similarities and differences between Australian and New Zealand trainee performance, and between annual cohort's performance, and summarise the findings.

Patterns of relationships between selection and assessment performance

Selection performance.

High mean scores in selection reflected highly competitive sets of candidates. Recognising that this data is generated from only those who were selected into surgical training, nevertheless, high scores and narrow standard deviations (SD) of the total selection scores (Total sel) indicate that most candidates performed strongly and were closely matched. The greatest Total sel SD occurred in 2008, for Australian (Au) candidates. This reflected a broader range of scores, particularly in the structured Curriculum Vitae (CV) and interviews, possibly revealing a disparity between scores (and performance) of ex-Basic Surgical Training (BST) candidates and 'novice' candidates, (2008 had the greatest proportion of applicants who had been BST trainees). Ex-BST candidates may have scored higher than others in the CV, which, in 2008 allocated more points for surgical experience. The other main area where the 2008 selection process differed from 2009 and 2010 was the interview, which in Au in 2008 was conducted at a single interview station. It is unclear how, or if this format might have affected the scoring, although other studies have shown that long form interviews tend to be less reliable than multi-station interviews.

The low mean scores for CVs in all years may represent stringent scoring by reviewers or may indicate that applicants, prior to applying to Surgical Education and Training (SET), were not gaining experiences in activities for which General Surgery (GS) allocated CV points. Possible reasons for this include: candidates applying at their earliest opportunity—before undertaking activities scored in the CV, or candidates being unsure of what activities were scored, or candidates having limited access to scored activities. As the CV had the least weighting of the selection tools it is possible for these candidates to have scored lower on the CVs than did candidates who were not selected into training—high scores in the RR and Int may have offset low CV scores of selected candidates. Harsh scoring by reviewers is unlikely, as scoring the structured CV required minimal subjective judgement.

The CV consistently revealed greater range of scores in candidates' performances than did the structured referees' reports (RR) or Int, as manifested by larger CV SDs for all cohorts. Candidates' high scores and closely matched performance in the RR and Int limit the ability of these tools to discern between candidates. The narrow RR and Int SD also effectively place greater emphasis on the CV as a ranking tool, despite its lower weighting (Eva & Reiter, 2004, p. 171).

Given the similarity of the Au and NZ selection tools, the alignment of performance between Au and NZ candidates in all selection items indicates that the candidates are closely matched between Au and NZ. The greatest difference in selection scores between the two countries occurred for the 2010 Int; this reflected not only a disparity between NZ and Au Int scores, but between NZ Int performance in 2010 and previous years' NZ Int performance. It is not clear why the NZ mean 2010 Int score was lower than the two previous years' scores.

The weak correlations and paucity of associations between performance in the selection items (i.e. between the CV, RR and Int), in all three years, reveal that candidate performance in any single selection tool was not analogous to their performance in either of the other selection tools. This implies that each of the selection instruments captured information about different attributes with few candidates demonstrating all of these attributes.

Weak, negative associations between the Au CV and Au Int seen in two cohorts (2009, 2010) reveal that candidates with higher scores in the Au CV were likely to receive lower scores in the Au Int. This trend supports the suggestion that low scores in the CV did not preclude candidates from high scores in the Int. The assertion that candidates who concentrated on enhancing their CVs did not demonstrate skills and attributes tested in the Int is also supported. It is possible that these candidates with high-scoring CVs had applied to SET unsuccessfully in previous years—gaining 'scored' experiences, but consistently scoring low for qualities that were valued in the Int. Future studies, noting performance and re-applications by non-selected candidates in successive cohorts could reveal whether this inference were supported. Alternatively, inverse associations in 2010 between performance in the Au Int and performance in both the Au CV and the Au RR, could indicate low predictive validity of the 2010 Au Int. This theory is supported by a lack of association between performance in the 2010 Au Int and any of the subsequent assessments. The increasing mean score for the RR over the three cohorts may indicate that candidates were performing clinically to a higher standard, or may indicate ‘score inflation’ whereby referees gave unwarrantedly high scores to facilitate selection of some candidates.

As noted in the Methods chapter (see Interview format p. 113 and Table 7 *Format of GS interviews 2008-2010*), the Au Int was changed from a long form interview to a multi-station interview in 2009, this coincided with the first instance of a negative association between the Au CV and the Au Int. In 2010 the Au Int was negatively associated with both the Au CV and the Au RR. The moderate, inverse relationship between the combined years’ NZ RR and the NZ Int implies that, although not significant in any single cohort, there was a trend over the three years for strong clinical performance, as manifested by the RR, not to be reflected in the Int.

The constancy of mean Int scores, which (with the exception of 2010 NZ; 73.16%) ranged from 80.30% (2008 NZ) to 83.63% (2009 Au), could indicate that most candidates were closely matched and performed at a high standard; that the interview tools were stable; that interviewers tended to score leniently; that interview standards were pitched too low; and/or that interviews did not discern well between top-performing candidates. This supports a hypothesis that the interview tools were reliable, but may have had limited effectiveness in measuring what they were intended to measure.

Assessment performance.

Examinations.

The set of strong associations in examination performance—spanning all examinations, all years and both countries—demonstrate that performance in all examinations was aligned, supporting assumptions that Au and NZ cohorts were comprised of trainees with similar knowledge and skills, that the examinations were of a consistent standard from one year to the next, and that the Generic surgical science examinations (GenSSE) and Specialty surgical science examinations (SpecSSE) tested similar knowledge.

The 2008 cohort achieved marginally lower mean scores in the Clinical Examination (CE) than did the 2009 and 2010 cohorts. A notable difference between the 2008 cohort and the subsequent two cohorts was that a higher proportion of the 2008 cohort had participated in the previous basic surgical training (BST) program (Collins et al., 2007). Some of this cohort first attempted the CE as BST trainees. As CE data are presented for candidates' first attempt at the examinations, this represents scores achieved by these trainees during BST. CE delivery parameters remained constant from BST to SET and it is not clear how BST participation could have influenced performance differences between the 2008 cohort and subsequent cohorts. Possibilities to explore include BST training not supporting trainees in the skills tested in this examination to the extent that SET training did, and that BST CE assessors may have scored more harshly than SET CE assessors.

Despite the CE differing from the SSE examinations in format, targetted skills and scoring methods, the strength of correlations between CE performance and SSE performance were only marginally lower than the strength of correlations between GenSSE and SpecSSE performance. Factors possibly contributing to the lower correlations between CE performance and GenSSE and SpecSSE performance include the delivery of the CE across multiple venues and the subjective nature of scoring the CE. Regional variances throughout Au and NZ may have decreased the uniformity of conducting and scoring the CE. Scoring CE performance was more subjective than scoring the machine-marked GenSSE and SpecSSE, placing greater reliance on the skills and judgement of individual assessors and on their knowledge of the required standards of performance. The moderate to strong associations between the CE and the SSEs indicate that the performance standards were aligned and that CE assessors understood the standards of performance required of candidates and consistently scored trainee performance to these standards.

The findings reveal that trainees with high scores in any single examination were likely to score high in all examinations—top trainees were likely to perform well in many facets of performance. This also supports the reliability of the examinations over several years, as examination performance data was drawn from results spanning roughly 10 years. These summative examinations, the most 'formal' of the RACS' SET assessments under review, have been subject to more rigorous analysis than were the workplace assessments. Regular, extensive standard-setting and review contributed to examination validity and reliability.

Direct observation of procedural skills (DOPS).

Each DOPS assessment represents an observed assessment of a trainee's skills in a particular procedure at a particular point in time. These assessments were intended to be formative—to provide feedback to the trainees so they might gauge their performance against a standard and take steps to improve if necessary. Conversion of qualitative DOPS assessments to numeric scores, while facilitating quantitative analysis for the purpose of this study, does not reflect the intention of this assessment.

DOPS mean scores were high—consistently over 78%. The DOPS SD tended to be large; the few scored items (n = 9) in DOPS assessments connoted that each item score assumed a considerable proportion of the total assessment score, increasing the SD. Au DOPS mean scores increased from DOPS1 to DOPS4 in all cohorts. This is consistent with trainees' skills improving during the usual 18-24 months between DOPS1 and DOPS4. NZ DOPS mean scores similarly increased from DOPS1 to DOPS2. The small number of NZ DOPS3 and DOPS4 reports limits the utility of performance data from these assessments; these are the only DOPS mean scores below 78%.

The 2008 NZ cohort mean scores for DOPS1 and DOPS2 were higher than the 2008 Au mean scores and higher than the combined years mean and the mean per year. This may indicate a tendency for 2008 NZ trainees to perform to a higher standard, or may imply that NZ supervisors were scoring more leniently than their Au counterparts. The DOPS were first implemented in 2008, and supervisors' unfamiliarity with these assessments, and inexperience in scoring them, could have reduced the reliability of their scores. Au and NZ scores were similar for the later cohorts' DOPS1 and DOPS2 assessments.

The moderate associations identified for the 2008 and 2009 cohorts between Au DOPS1 and Au DOPS2, highlight a consistency of procedural performance and/or assessment during the first year of SET training. This result, and the high mean scores noted for these assessments, may reflect that trainees are well prepared for these assessments, that the procedural skills undertaken by trainees during this early stage of training were the least complex and may have been the least challenging and/or reflected a lenient approach by assessors during the early stages of training.

When the three cohorts' data were combined, the associations between Au DOPS1 and Au DOPS2 were endorsed and associations were revealed between DOPS2 and DOPS3. Associations for the combined years' performance in DOPS2 and DOPS3 indicate that although not significant for any individual year-cohort, there was a tendency for trainees' procedural performance to be consistent from one year to the next.

Mini clinical evaluation exercise (MiniCEX).

Each MiniCEX assessment represents an observed assessment of a trainee's skills in a particular clinical examination at a particular point in time. These assessments were intended to be formative—to provide feedback to the trainees so they might gauge their performance against a standard and take steps to improve if necessary. Conversion of qualitative MiniCEX assessments to numeric scores, while facilitating quantitative analysis for the purpose of this study, does not reflect the intention of this assessment.

MiniCEX mean scores were high—consistently over 81%. The MiniCEX SD tended to be large; the few scored items (n = 10) in MiniCEX assessments connoted that each item score assumed a considerable proportion of the total assessment score, increasing the SD. Au and NZ MiniCEX mean scores tended to increase from MiniCEX1 to MiniCEX3, although not as consistently as noted for the DOPS assessments. NZ MiniCEX mean scores in all cohorts increased from MiniCEX1 to MiniCEX2. The small number of NZ DOPS3 and DOPS4 reports limits the utility of performance data from these assessments.

The lack of associations for MiniCEX performance among the 2008 and 2010 cohorts reveals a lack of consistency of clinical performance and/or assessment, in contrast to the 2009 cohort's consistency of performance in the Au MiniCEX1 and MiniCEX3. The combined cohorts' data confirmed the relationship between Au MiniCEX1 and MiniCEX3 and revealed a correlation between Au MiniCEX3 and Au MiniCEX4 that, although not significant for any individual cohort, may indicate a tendency for Au trainees' clinical performance to be consistent during the second year of training.

In comparison to the DOPS reports, the dearth of significant correlations among the MiniCEX reports may reflect the unpredictability of clinical encounters. Trainees nominated the procedures in which they were assessed by DOPS and could therefore anticipate requirements, however the MiniCEX assessments were conducted during incidental clinical encounters for which trainees could not specifically prepare. Particularly during the early SET years, trainees may have been challenged by unfamiliar circumstances and patient conditions. The unpredictability of patient presentations combined with limited diagnostic and patient-management experience may have resulted in trainees' inconsistent clinical performances.

End of term assessment (ETA).

Each ETA assessment represents observed appraisals of a trainee's skills during a six-month clinical rotation. These assessments were intended to be primarily formative, with some summative capacity, as identified in the Methods chapter (p. 133). Conversion of qualitative ETA assessments to numeric scores, while facilitating quantitative analysis for the purpose of this study, does not reflect the intention of this assessment.

Mean ETA scores were high—consistently over 83%; (with two exceptions: 2009 NZ ETA3 (79.30%) and 2010 Au ETA1 (82.66%)). The SD tended to be moderate, reflecting the large number of scored items (n = 40 items) and narrow range of scores achieved by trainees in

ETA assessments. The constancy of mean ETA scores, which (excepting 2009 NZ ETA3 and 2010 Au ETA1 noted above) ranged from 83.18% (2009 NZ) to 87.73% (NZ 2008), could indicate some, or all of the following: that most trainees were closely matched and performed above the required competency standards; that changes to the assessment tools were minimal; that assessors tended to be lenient; that assessment standards were pitched too low; that these assessments did not discern well between high-performing trainees. This supports a hypothesis that these assessments were reliable, but may have had limited effectiveness in measuring what they were intended to measure.

The ETAs, unlike the CE, DOPS or MiniCEX assessments, summarised observed trainee performance in all clinical activities during the 6-month rotation with items in the ETAs presenting a more nuanced assessment of the generic nine RACS competencies. The longitudinal nature of the ETA and the comprehensive scope of the items were likely to preclude trainees from preparing for specific ETAs; rather, trainees who directed their ongoing preparation—across the range of competencies—to anticipated daily clinical activities were most likely to score high in ETAs.

The 2009 cohort presented more associations for ETA performance than did the 2008 cohort; performance of the 2010 cohort is partly unknown due to incomplete data. The 2009 correlations between ETA1 and ETA2, ETA1 and ETA3, ETA2 and ETA4 indicate a tendency for trainees' clinical performance to be consistent from one year to the next as they progressed through SET.

When data for the three cohorts was combined, additional associations were identified; associations were discerned between ETA1 and all subsequent ETA assessments, between ETA2 and all other ETA assessments, between ETA3 and all other assessments, and between ETA4 and all other assessments. This supports the hypothesis that trainee clinical performance was consistent during the first two years of training. It also implies that ETA assessors understood the standards of performance required of candidates and consistently scored trainee performance to these standards. The findings reveal that trainees with high scores in one ETA were likely to score high in all ETAs—top trainees were likely to perform well in many facets of performance.

Comparisons of trainee performance in selection and in subsequent assessments.

Pearson correlation analyses—comparing selection with assessments.

Pearson correlations analyses revealed associations between selection and assessments. The 186 associations discerned between scores in selection and scores in assessments during training represented approximately a quarter of the 776 associations possible, with most of these associations being weak (n = 100) or moderate (n = 40), revealing limited links between performance in selection and performance during surgical training. More associations were identified between selection and assessments for Au trainees (n = 63) than for NZ trainees (n = 51), and the greatest number of associations were discerned when the two countries' data were combined (n = 72). These results must be considered in the knowledge that data were held for 64 fewer NZ assessments than Au assessments. See Results chapter Table 15 and Table 16. Assessment methods, items and content were consistent between Au and NZ, therefore correlation and regression differences identified between the two countries may indicate differences between their selection processes and/or differences between candidates and/or primary medical training and/or specialty medical training practices.

The tendency for RR, Int and Total sel scores to be positively correlated with assessments highlights consistency of trainee performance in selection and during training and alignment of these selection items to assessments. The predominantly negative correlations between the CV and assessments reveal that applicants who scored high in the CV were less likely to score high in assessments, indicating that items scored in the CV were not aligned with SET assessments. The negative association of the CV with SET assessments may reflect differences between components of the CV—scoring professional, academic and personal achievements—and attributes required to succeed in the early years of surgical training. The CV is the only selection tool in which candidates could purposefully organise their activities to maximize their score, however a high score for presentations, publications and prizes, for instance, may not presage knowledge and skills required in examinations or in clinical contexts.

The Total sel, as the score by which candidates were ranked and admitted to SET, was intended to reflect candidate performance in the CV, RR and Int, and the proportional weightings of these three selection items. Despite the proportional weightings of the selection items, the number of significant correlations for the Total sel (n = 43) was closer to those observed for the CV (n = 40) than for the preponderant RR (n = 56) (see Results chapter Figure 15). The influence of the CV on the Total sel may have been unintentionally elevated due to the CV SD being greater than the RR SD and the Int SD. The broader range of scores in the CV denoted that it discerned better between candidates than did the RR or Int and resulted in the CV having greater influence on the Total sel than its weighting would imply.

Most significant correlations were weak to moderate, with roughly 35 percent of these being negative and 65 percent being positive associations. Most significant correlations between selection and assessments were identified for the 2009 cohort, although this must be understood in the context that data for the 2010 cohort's assessments were incomplete. The combined years' data, although not identifying as many significant correlations as the 2009 cohort, disclosed more significant correlations than either the 2008 or the 2010 cohorts; the 2008 cohort exhibited the fewest associations. The significant correlations identified for the 2009 cohort were not only more numerous, but also tended to be stronger than those from 2008 and 2010 (see Results chapter Figure 23). These findings are greatly influenced by significant correlations for the 2009 RR and Total sel with many assessments (particularly with the ETAs) and by significant correlations for the 2009 Int with the three examinations. Although the RR demonstrated the greatest number of significant correlations, the Int tended to be more strongly associated with assessments (see Results chapter Figure 24).

Assessment items that were identified with 10 or more significant correlations across the three yearly cohorts were the CE (n = 19), DOPS1 (n = 15), ETA1 (n = 15), ETA4 (n = 18) and Av ETA (n = 17), all of these—except DOPS1—could be termed 'in-depth' assessments of clinical skills, highlighting an alignment of selection with practical, clinical components of SET, a proposal also supported by the predominance of positive associations between selection and these assessments. The only occurrence of significant negative correlations outnumbering significant positive correlations for any assessment item was for DOPS2 (see Results chapter Figure 29). As the DOPS assessments are a comparatively cursory and superficial snapshot of performance, individually they tend to be less rigorous than the CE or ETAs.

The strongest significant correlations with assessments were identified for NZ cohorts, (21 strong and four very strong correlations); this compares with one strong and no very strong correlations between selection and assessments identified for Au cohorts. The weakest significant correlations with assessments were identified for Au cohorts, (45 weak and seven very weak correlations); this compares with nine weak and no very weak correlations between selection and assessments identified for NZ cohorts (see Results chapter Figure 22). These differences might be attributed to the smaller number of NZ than Au trainees, or other differences between the two countries' selection, training and assessment practices. Although the SET curriculum is identical in both countries, implementation differences affecting assessment results might have occurred; the smaller number of NZ trainees might have resulted in individual NZ trainees receiving more, or more personalised training than their Au counterparts. These considerations could be explored in future studies.

The alignment of the RRs with ETAs reflects similarities between these assessments; both evaluated clinical performance in similar competencies over time. Like the ETAs, the structured RRs required referees to rate candidates against performance items in each of the nine RACS competencies.

Regression analyses—Comparing selection with assessments.

Regression analysis revealed fewer associations between selection and assessment performance than did Pearson correlations. Both forms of analysis identified the RR with most associations and the CV fewest. No individual selection item strongly predicted performance in all assessments. The CV was aligned with performance in the GenSSE and one DOPS assessment and was negatively associated with performance in eight assessments (CE, DOPS, Av MiniCEX and ETA3); RR scores anticipated future performance in 36 assessments (SpecSSE, CE, DOPS, MiniCEX and ETAs;) the Int anticipated performance in 18 assessments (GenSSE, SpecSSE, CE and ETAs, and one DOPS) Total sel anticipated performance in 31 assessments (GenSSE, SpecSSE, CE, DOPS and ETAs). These findings support the conclusion that attributes and behaviours assessed by the RR were most consistently aligned with attributes and behaviours assessed during surgical training., whereas those assessed by the CV were the least consistently aligned with these attributes and behaviours.

Yearly cohorts. Results of regression analyses were consistent with Pearson correlation analyses for yearly cohorts. The 2009 cohort and combined years' performance data indicated more alignment of performance in selection with performance in training than did the 2008 or 2010 cohorts.

Countries. Results of regression analyses per country were also consistent with associations identified by Pearson correlations. The greatest number of significant regressions were identified when the two countries' data were combined.

In-training assessments—examinations. No significant regressions were detected between selection scores and examination scores for the 2008 cohort, however the 2009 and 2010 NZ Ints were consistently associated with examination performance, with stronger associations evident in 2010. The 2009 Au Int was also associated with all examinations; this relationship extended to the CE but not the written examinations in 2010. The 2009 Au RR, Int and Total sel performance predicted CE scores. The 2009 Au Int and 2009 NZ Int were both associated with all three examinations (see Results chapter Table 17); this is the only occasion in which the Au and NZ results were so consistent.

Components of 2010 NZ selection performances were predictive of scores in all three examinations—the 2010 NZ Int and Total sel were aligned with performance in the Generic and SpecSSE and the 2010 NZ CV was negatively associated with the CE scores. However, no 2010 Au selection item predicted examination performance (see Results chapter Table 22). Multiple regression analysis of the combined years' data confirms that the NZ Int and NZ Total sel scores, reflecting 2009 and 2010 outcomes, predicted performance in the two written examinations (see Results chapter Table 23). The NZ CV also was aligned with GenSSE scores—one of the few instances of the CV being positively associated with assessments. Au (and ANZ) combined years' data identified that the Au RR, Int and Total sel predicted CE performance, although no NZ selection item predicted performance in the CE (see Results chapter Table 23).

In-training assessments—work-based assessments. Performance data for 2008 and 2009 showed few alignments between selection and DOPS assessments, however the 2010 Au cohort had predictive relationships for CV, RR and Total sel with DOPS1, and a negative predictive relationship for Int with DOPS3. Combined years' data also showed relationships for RR and Total sel with DOPS1, and an inverse relationship for NZ CV with Average DOPS. Few predictive relationships between selection items and DOPS assessments were identified with no discernible pattern of selection items, years, or countries. The 2010 Au Int accounted for 19.8% of the variance in DOPS3, the relationship being negative, indicating that high scores in the 2010 Au Int were likely to be associated with low scores in DOPS3.

There were fewer associations between selection and MiniCEX performance than observed for DOPS performance, attesting that there is less alignment between selection and MiniCEX than between selection and DOPS. The only predictive relationships identified for MiniCEX assessments were for Average MiniCEX scores. These were identified for the 2009 Au cohort, and NZ combined years. No predictive relationships were identified through regression analysis for the 2008 or 2010 cohorts, nor for any individual MiniCEX assessment. The RR was positively aligned with Average MiniCEX and the CV negatively, but neither the Int nor Total sel was aligned with MiniCEX performance.

Selection scores were most frequently aligned with ETA performance, particularly for the 2009 cohorts. The combined years' data showed that selection predicted ETA3 and ETA4, and the combined years' selection scores also predicted the average ETA scores for both countries.

The inverse association of the 2008 Au Int with ETA2 was a curiosity (see Results chapter Table 35 and Table 36). There was little synchronicity between Au and NZ regarding the degree of significance of selection items predicting ETA performance (see Appendix A Table 70).

The 2009 cohort showed many associations between performance in selection and performance in ETAs. The 2009 selection item most frequently associated with ETA performance was the RR, the 2009 Au RR being associated with all four ETA assessments and the 2009 NZ RR being associated with two ETAs (ETA2 and ETA4). The 2009 NZ Int was also associated with two ETAs (ETA1 and ETA3), although the 2009 Au Int was not associated with any ETA scores. As noted for the 2008 cohort, neither the 2009 Au CV nor the 2009 NZ CV predicted ETA performance. Unsurprisingly, associations were found between both countries' 2009 RRs, the 2009 Au Int and both countries' 2009 Total sel and the Average ETA scores.

Multiple regression analysis confirmed the findings of association: the 2009 Au RR predicted performance in all Au ETA assessments, while 2009 NZ RR predicted performance in ETA2 and ETA4 (see Results chapter Table 31). This is the only instance identified in which a selection item (the RR) predicted performance in all four assessments. The 2009 Au Total sel, influenced by the 2009 Au RR predicted performance in ETA1 and ETA2, while the 2009 NZ Total sel, unsurprisingly, predicted ETA2 and ETA4. Interestingly, the strongest association—between 2009 NZ Int, and ETA3—did not resolve as a predictive relationship.

The capacity of both the Au and NZ RR as predictors of ETA performance was evident, with neither the CV nor the Int predicting Au or NZ ETA performance; however, when the Au and NZ scores were combined, the Int predicted ETA1 and ETA2 performance, as well as Average ETA performance. The 2009 Au and NZ cohorts showed more similarities in performance than was seen for the 2008 and 2010 cohorts.

2010 selection scores compared with ETA scores. As seen in the 2009 cohort, 2010 Au RR and 2010 Au Total sel were associated with ETA1, however, the 2010 cohort's

ETA1 associations did not continue to ETA2. Data for this cohort was limited to the first two ETA assessments. In contrast to the 2009 cohort, none of the 2010 NZ cohort's selection items predicted ETA performance.

Further study, including the ETA3 and ETA4 results would indicate whether the many associations seen in the 2009 cohort were an anomaly or whether the 2010 cohort also showed these associations. Such analysis might also reveal patterns of association for the 2010 cohort from the early to the later ETA assessments.

Selection vs ETAs. Combined years. When the three cohorts' data were combined, the Au RR was associated with performance in three ETAs, and predicted performance in ETA3 and in ETA4. The CV was not strongly associated with ETA performance—the Au CV was only associated with performance in ETA3, this negative association also predicting performance in this assessment; the NZ CV was not associated with ETA performance at all. The Au Int did not predict ETA performance, however the NZ Int predicted ETA3 and the Average ETA performance. The combined cohorts' data support the inference that selection items are associated with performance in the later ETA assessments and that the differences between the countries' implementation of the selection instruments or the ETAs affects the interrelationships. The Au CV and Au RR are associated with performance in ETA3 and ETA4, with the NZ RR and NZ Int also being associated with performance in these ETA assessments.

Similarities and differences between Australian and New Zealand performance in selection and assessment in the bi-national GS SET program

The RACS' SET program was common to both Au and NZ. Differences between Au and NZ's populations were reflected in the proportions of Au and NZ GS trainees. Au and NZ each administered selection to GS locally and had some autonomy regarding the content of selection items. Examinations and work-based assessments (WBAs) were identical for both countries. The current study reviewed Au and NZ selection and assessment practices, but did not consider societal differences between Au and NZ populations or surgical trainees. I propose that RACS, GSA and NZAGS assumed that the Au and NZ populations were similar enough that shared selection and assessment practices would neither favour nor discriminate against either population. Similarities and differences between Australian and New Zealand surgical selection and assessment performance outcomes are of interest to the current study because they could confirm assumptions about parallel practices and similar trainee populations or could denote previously unacknowledged differences in the content and implementation of Au and NZ selection or assessment items, or of training practices, or between the Au and NZ trainee populations; differences could warrant further review of relevant aspects of the training program.

Selection performance—similarities and differences.

As could be expected based on relative populations, approximately six times as many trainees were selected into the Au GS training program than into the NZ training program. Generally, both countries' mean scores for CV, RR, and Total Sel reflected a high level of similarity. With regard to Int score comparisons, one cohort score was appreciably different between the two countries—the 2010 NZ Int was noticeably lower than the 2010 Au Int score (see Appendix A Table 37). Given the similarities between each country's candidates in other selection components, the different 2010 Int outcomes most likely reflected the different question content and scoring mechanisms, unique to this cohort. The provision of statements of ideal responses and scoring guidelines as well as the differences in the specificity and standards of knowledge between the Au checklist and NZ 'issues' statements could have affected scoring in the 2010 Int. Overall however, Au and NZ candidates performed similarly and it appears that consistency in tools is a key factor in the observed similarities in bi-national selection performance.

Assessment performance—similarities and differences.

In general, mean scores in examinations and ETAs were analogous for the two countries, however, bi-national similarity was less prominent in DOPS and MiniCEX performance. The 2008 cohort's mean DOPS1 and DOPS2 scores (DOPS1: Au = 79.78%, NZ = 84.76%; DOPS2:

Au = 78.59%, NZ = 85.37%) implied that these Au trainees performed at a lower standard than NZ trainees or that Au assessors may have been harsher in the early DOPS in that year (see Appendix A Table 39). The 2008 Au DOPS1 and DOPS2 scores were also lower than later DOPS scores (DOPS3: Au = 85.78%; DOPS4: Au = 84.16%), possibly indicating acquisition of procedural skills as training progressed. Although the two countries' average MiniCEX1 and MiniCEX2 scores were alike for the 2008 cohort, for 2009 and 2010 cohorts, the Au average MiniCEX1 and MiniCEX2 scores tended to be lower than mean NZ MiniCEX scores (see Appendix A Table 40). This may be an artefact of low NZ trainee numbers (2009: MiniCEX1 NZ n = 10; MiniCEX2 NZ n = 8; 2010: MiniCEX1 NZ n = 17; MiniCEX2: NZ n = 8), although inconsistencies in scoring may have resulted from variations in assessors' interpretations of the expected performance standards of the trainees or of the processes, tools and purpose of these work-based assessments; also, trainees in the very early stages of training would have had differing exposure to clinical experiences and likely had not built a comprehensive repertoire of skills. In particular, in the MiniCEX assessments trainees may have been assessed on diverse clinical encounters that had varying levels of difficulty, or that they were not yet familiar with. These differences are unlikely to be specific to either country, however the impact of inconsistencies of trainee experience and variability in assessor scoring may have reduced the reliability of these assessments, thereby reducing the likelihood of alignment between countries. The small number of assessments undertaken per trainee may also have reduced the reliability of the MiniCEX assessments; the two MiniCEX assessments per year undertaken by GS trainees being fewer than the four to ten recommended by **Norcini** (2003, 2005) in his studies of internal medicine residents' clinical skills.

As Au and NZ conducted identical examinations with standardised scoring and trained assessors, similarities in examination performance were expected and corroborate assumptions that trainees from the two countries had similar professional, academic and societal backgrounds. In contrast, the work-based assessments—DOPS, MiniCEX and ETAs—implemented as individual one-to-one appraisals of diverse activities and conducted by numerous supervisors with varied expertise in training and assessment presented more opportunities for variation in scoring outcomes. Compounding this, the DOPS and MiniCEX were newly introduced in 2008 and training in their use was limited. Supervisors and trainees were more familiar with the ETAs, for which Au and NZ scores were mostly aligned—the noticeably lower scores for the 2009 NZ ETA3 assessment (see Appendix A Table 41), was based on a small number of NZ trainees (n = 7) and the data may have been skewed by one or two low scoring NZ trainees.

These findings support assumptions that Au and NZ GS trainees were alike, received similar training experiences and performed alike in most assessments. Where the stringency or content of assessments differed, trainee performance outcomes also tended to differ. No major national differences were identified in the implementation of any SET assessments; the lack of alignment between Au and NZ DOPS and MiniCEX performance is therefore more likely attributable to the 'generic' GS implementation of the assessments than to national differences between trainees or training programs. RACS and GS may consider reviewing work-based assessment practices and assessor training to increase the reliability of outcomes.

Similarities and differences between Au and NZ selection and training relationships.

Most of the identified relationships between selection and assessments were positive for both countries (see Results chapter Figure 28), indicating bi-national alignment of selection and assessment performance. Although more significant correlations and predictive relationships were identified for Au than for NZ, the NZ relationships tended to be stronger (see Results chapter Figure 22). These differences could have been influenced by discrepancies in data available for each country, as more Au than NZ assessment results were considered; numbers of trainees per country may also have affected the correlation outcomes. Most predictive relationships for both countries were identified between selection and the ETAs (see Results chapter Figure 30). The CV was a poor predictor of subsequent performance for both countries—the predominantly negative associations identified between the CV and assessments highlighting a shared disparity between experiences scored in the CV and those valued in SET.

The RR was the strongest predictor in Au selection, while the Int was the NZ selection item with the greatest number of predictive relationships. Total sel, as the second most frequent predictor of assessment performance for both countries, indicates that despite national differences in the predictive ability of individual selection tools, the scores that were adopted to rank candidates for selection were, in general, aligned with performance in training—particularly as there were no inverse predictive relationships from the Total sel scores (see Results chapter Figure 31). When the two countries' data were combined, more significant correlations (ANZ n = 72) and predictive relationships (ANZ n = 38) were identified than were identified for either country individually, demonstrating that tendencies that were not significant for one country were augmented in the combined group.

The differences between the two countries' 2010 Int questions identified previously, resulted in different predictive outcomes for the Int. The 2010 Au Int scores were associated with CE performance and inversely predicted performance in one DOPS, while the 2010 NZ Int predicted performance in the two written examinations (the GenSSE and the SpecSSE). This indicates that the 2010 NZ Int was more strongly aligned with assessments, particularly the written examinations, than was the 2010 Au Int. This represents a rare instance in the body of research in selection to surgical training of direct comparisons of one facet of selection being possible in an actual high stakes selection process (for example: Bann & Darzi, 2005; Kenny, McInnes & Singh, 2013; Maan et al., 2012; Patterson et al., 2009; Prager et al, 2010; Siu & Reiter, 2009).

Au selection scores predicted performance in assessments of clinical skills (CE, DOPS and ETAs) more times than did NZ selection scores. Five NZ selection scores predicted performance in the written examinations contrasting with Au results, in which no selection item predicted performance in these examinations, (see Results chapter Figure 30). Overall, Au selection outcomes appear to reflect the clinical aspects of training while NZ scores are more representative of the cognitive aspects, as manifested in the written examinations.

These results support assumptions of homogeneity between Au and NZ candidates (in similar situations, under similar conditions they were likely to perform alike), in consistent implementation of selection and assessments between the two countries and in the use of Total sel as the final arbiter of selection. They also indicate that further review of the content and implementation of the CV, DOPS and MiniCEX could be undertaken to increase the alignment of selection (particularly Total sel) with training requirements and to increase the DOPS and MiniCEX's capacity to accurately reflect trainee standards of performance. It may also be appropriate to further explore why Au relationships between selection and assessments tended to be weaker than NZ relationships.

Patterns of relationships between yearly cohorts

This section will discuss differences in the significant correlations and predictive relationships that were identified between yearly cohorts.

Selection performance—similarities and differences.

Although trainee intake was stable in 2008 and 2009 the increase in the number of Au trainees in 2010 was marked (see Results chapter Table 13). Reasons for this increase were not explored in the current study. As no extreme performance differences were observed between the 2010 cohort and previous cohorts the increased intake appears to have had little effect on training outcomes.

The similarity of mean CV scores in 2008 and 2010 supports assumptions that candidates had similar experiences prior to applying to SET in these years and scoring allocations were similar. Lower CV scores in 2009 may indicate that candidates were less experienced, or may demonstrate subjective influences on CV scoring or may reflect that the score allocations within the CV differed in 2009—allocating fewer points for surgical and medical experience and increasing points allocated for qualifications. As identified in the Methods chapter, scoring of the CV components did change for each cohort in the study (see Methods chapter Table 4). The 2009 CV content and component scoring differed from the other two cohorts, placing more emphasis on research, publications and presentations and less weighting on surgical and medical experiences. The 2008 and 2010 CV components were similar to each other, with only slight differences in maximum allowable component scores. Mean RR scores markedly increased across the three cohorts, possibly signifying that referees were becoming more lenient over the years or that later candidates were performing to a higher standard and therefore might be better prepared for surgical training; the content of the RRs was identical for the 2008 and 2009 cohorts, but for the 2010 cohort, although the sections within the RR remained constant, the questions and descriptors within the sections changed. The differences in the associations found for the 2008 RR and 2009 RR—with the 2009 RR demonstrating the greatest number of significant correlations (see Results chapter Table 16)—were unanticipated and could result from influences such as differences between trainee cohorts, between referee assessments, or these cohorts may have been atypical. Further studies would be required to discern the reasons for this outcome. Mean Int scores were consistent for the three cohorts with the exception of the 2010 NZ Int as noted above. The 2010 NZ Int scores were not only lower than 2010 Au scores, but were also lower than previous NZ cohorts. This may have resulted if NZ Int questions were pitched at a higher standard of knowledge or performance, or if scenarios were less relevant to candidates or if the 'issues' statements were broader or more restrictive than in previous years. As other Int mean scores were relatively stable, the tendency for Total sel scores to increase across the three cohorts predominantly reflects RR performance. Key distinctions between the three cohorts' selection scores were the 2008 cohort tending to score lower in the RR and Total sel and the 2009 cohort scoring low in the CV; the 2010 cohort scored highest for the RR and Total sel. This may reflect 'score inflation' in scoring the RRs, whereby referees tended to increase scores for subsequent cohorts (see Appendix A Table 37).

Assessment performance—similarities and differences.

Mean scores in all examinations were stable across the three cohorts except for the 2008 cohort's CE results, which were somewhat lower (2.6%-5.1% lower) than scores achieved by the later cohorts. As trainees could choose when they undertook the CE, lower scores for the

2008 cohort are more likely to highlight that these trainees performed at a lower standard than any instance of the CE being assessed more, or less harshly than any other. It is possible that trainees in later cohorts received better preparation for this examination during training, or, consistent with the higher RR scores, that later cohorts performed to higher standards in CEs. Further study of subsequent cohorts' performance in the CE may reveal whether the 2008 result was an anomaly (see Appendix A Table 38). Average scores in the work-based assessments—DOPS, MiniCEX and ETAs—were relatively stable across the three cohorts, although individual DOPS and MiniCEX scores fluctuated. Overall, trainees in the three cohorts performed alike in procedural skills and clinical activities during early surgical training.

In summary, scores in the written examinations were moderately consistent across cohorts, however, the 2008 cohort tended to score lower in the CE. DOPS and MiniCEX scores varied within and between cohorts and average ETA scores were stable across cohorts (see Appendix A Table 39, Table 40 and Table 41).

Similarities and differences between 2008, 2009 and 2010 cohorts' selection and training relationships.

The fewest significant correlations between selection and assessment performance were found for the 2008 cohort and the most were identified for the 2009 cohort. As the 2010 cohorts' data were incomplete the 2010 results are inconclusive however, based on performance data available, it is unlikely that complete data for the 2010 cohort could equal the number of significant relationships identified for the 2009 cohort. Patterns of association varied between cohorts: roughly half of the significant correlations for the 2008 cohort were inverse relationships between the CV and subsequent assessments, highlighting that the 2008 CV was a very poor predictor of subsequent performance; 2008 was the only year in which more negative than positive correlations were identified, corroborating the lack of alignment between selection and this cohort's subsequent performance. With regard to RR comparisons, only one 2008 Au RR score correlated with later performance (see Results chapter Table 16); in contrast, the 2009 and 2010 Au RRs were strong predictors of later performance. This counterintuitive result may warrant further study, as the 2008 and 2009 GS RRs were identical, but the RR was changed in 2010. One inference is that factors other than the content of the RRs affect selection–assessment performance relationships. The number of associations from the RR, the Int and Total sel peaked in 2009, primarily because for that cohort, the Au RR, and the NZ Int were associated with many subsequent assessments (see Results chapter Table 15, Table 16 and Figure 25). Numbers of inverse correlations decreased from 2008 to 2010, implying that selection tools and performance—particularly in the CV—became better aligned with assessment performance over the three cohorts.

Predictive relationships followed similar patterns to those identified for significant correlations—fewest were identified for the 2008 cohort and most were seen in the 2009 cohort. The current study did not identify reasons for the observed differences between the 2008 and 2009 cohorts' performance. Further study may reveal whether the number and pattern of associations identified for the 2009 cohort were an anomaly or whether the 2010 and subsequent cohorts displayed similar associations.

Summary of findings.

This study has noted that the selection items changed yearly, that the examinations were the most standardised of the assessments, that two WBAs (DOPS and MiniCEX) were newly introduced in 2008 and that selection items differed from most assessment items in content and implementation, (the exception being the similarity between the RRs and the ETAs). Annual changes to selection items restrict the ability to make meaningful comparisons between different cohorts' selection performance or to identify patterns of performance relationships between selection items and later assessments. The examinations, as the most rigorously implemented, monitored, and reviewed of the assessments, offer reliable parameters by which to measure the effectiveness of selection tools in cognitive and, to a lesser extent, affective domains. New WBAs and assessors with limited experience and training in their use may have resulted in variable scoring and inconsistencies between the clinical activities being assessed; these factors, particularly when combined with fewer than optimal implementations of the WBAs are likely to have decreased their reliability.

The most consistent finding from this study, spanning all years and both countries was alignment of RR performance and ETA performance—reflecting not only applicants' and trainees' ongoing, enduring clinical performance in the workplace, but also a consistency of content and methods used in these assessment instruments. It is notable that applicants to GS were not able to directly choose the consultants who completed the RRs, a factor that undoubtedly contributed to the impartiality and reliability of RR assessments. Associations between Int performance and ETA performance were also prominently represented in two cohorts (2008 and 2009), primarily for NZ; the combined influence of this and the RR on Total sel scores resulted in associations between Total sel and ETA performance—indicating Total sel as a useful predictor of later clinical performance. The few CV associations with subsequent assessments (examinations and ETAs) were consistently inverse, revealing the CV as a poor predictor of subsequent performance in GS surgical training. The Int was aligned with examination performance for two cohorts (2009 and 2010). Of the three annual intakes, the 2009 cohort demonstrated most associations between selection and assessments during training. Annual changes to selection instruments and low reliability in some assessments have no doubt contributed to the variable relationships between selection and assessment performance. Where selection instruments and assessments were consistent, with identified standards (for example, the examinations), there was greater similarity in associations identified bi-nationally and across yearly cohorts.

Other predictive validity studies have also found limited predictive validity of selection instruments on performance during training (Patterson et al, 2016; Roberts et al, 2017; Schaverien, 2016).

Part 2 Contextualising the Royal Australasian College of Surgeons' (RACS) surgical training practices

As the only accredited provider of training in nine surgical specialties for Australia (Au) and New Zealand (NZ), the frame of reference for the Royal Australasian College of Surgeons' (RACS) practices is global. In this section I discuss RACS' unique approach to surgical education and training, focussing on selection and assessment activities. I also reconsider themes introduced in the Literature Review, in relation to RACS' training program.

Organisational and international contexts for ANZ medical specialty selection, training and assessment

Organisational context of RACS' training.

From RACS' founding in 1927, (Beasley, 2002; "About RACS | Royal Australasian College of Surgeons", 2017), and contemporaneously with many international surgical training organisations, Australian and New Zealand (ANZ) surgical training evolved from the "chaotic mix of institutions and practices" criticised by Geffen (2014, p.19) and others (Beasley, 2002; Storey, 2014). The current study verified that in 2008, 2009 and 2010, ANZ General Surgery (GS) selection, training and assessment practices conformed to regulated processes, which contrasted with prior ANZ surgical training models and with many contemporary international models. From apprenticeships to competency-based training, diverse vying influences have shaped and continue to shape surgical training. Approaches to surgical selection, training, and assessment may juxtapose many of the attributes identified in the literature review, including: regulation and independence, accountability and ambiguity, egalitarianism and privilege, individuality and society, theory and practicality, objectivity and subjectivity, systematic and haphazard activities, planning and spontaneity, training and service (employment), rigidity and flexibility, uniformity and inconsistency, specificity and imprecision, criterion-referenced and ambiguous standards. Accountability, objective measurement and consistency of training and certification became priorities as surgical training worldwide coalesced around models that combined university and hospital-based activities. Despite this, in several countries, arbitrary and subjective selection arrangements continue to coexist with more objective approaches (Adam et al., 2011; Gough & Bell, 1989; Hern, Alter, Will, Snoey & Simon, 2013; Robins, McInnes & Esmail, 2014).

International context: Comparisons to other medical specialty selection, training and assessment practices.

Selection practices for primary and specialty medical training are shaped in part by custom and cultural contexts. Although many medical training programs have evolved from British, European or American models, variations in approaches within and between countries are widespread and a profusion of selection instruments and processes abounds. The vast majority of selection processes presume to differentiate between candidates in order to admit those who are most likely to perform at the highest standards during training and beyond (Carroll et al., 2009; Cleland et al., 2012; Cuschieri et al., 2001; Elfenbein et al., 2015; Gallagher et al., 2008; Martin, 1996). Commonly used selection instruments include measures of academic achievement and/or personal, or 'character' qualities, letters of recommendation, personal statements, references, interviews, and tests of practical skills, aptitude, psychometric ability, problem-solving, judgement and/or knowledge (Maan et al., 2012; Patterson et al., 2013; Roberts et al., 2017; Schaverien, 2016). Unacknowledged influences on selection could conceivably include the use of discretionary judgements by selectors who override formal selection outcomes, informal discussions about candidates or training positions between selectors and people outside the selection process and even inducements offered to selectors by candidates or their representatives.

In the manifold medical programs worldwide, selection instruments may be systematically or inconsistently implemented, may include subjective and/or objective measures and may be validated or unsubstantiated. Scoring may be criterion-based, norm-based or haphazard and imprecise. Selection to specialty training in many countries and programs involves processes that combine candidate performance in two or more selection instruments (also known as selection 'tools'). Individual programs stipulate the instruments, combinations and performance measures that define admission to their programs.

Even where surgical training programs share goals and practices, differences in implementation abound. Procedural fairness has been advocated, and explicitly prioritised in some surgical selection processes (Gallagher et al., 2014; Patterson et al., 2011; Thomas, Davison, Gee, Grant & Taylor, 2013) however, the profusion of selection methods and variability in the rigour with which they are implemented can compromise effectiveness in this domain. MMIs and SJTs have been shown to predict clinical performance (Eva, Reiter, Rosenfeld, Trinh, Wood, & Norman, 2012; Patterson et al., 2016). 'Fairness' may be limited to adherence to published protocols while implementing selection tools that may have little reliability or predictive validity. Additionally, unconscious biases in tools, procedures or implementation may unintentionally discriminate against particular groups or individuals. RACS' selection, training and assessment processes are conducted against this messy, dynamic global context, in which many participants are striving to establish and implement ordered, reliable systems and structures, by refining existing instruments and protocols and by creating new ones.

Australian and New Zealand (ANZ) surgical training—the Royal Australasian College of Surgeons' (RACS) surgical education and training (SET) program.

During the period of the current study, ANZ selection to primary medical training followed similar processes to international models—direct entry medical courses considered secondary school results and performance in the Undergraduate Medicine and Health Sciences Admission Test (UMAT); graduate entry programs considered one or more of: performance in the GAMSAT, grade point average (GPA) in a qualifying degree, performance in a portfolio, special application, supplementary form, autobiographical statement and performance in an interview (Graduate Entry Medical Schools Admission System, 2015). In the years under review, selection to ANZ surgical training differed from international training programs. RACS' SET was streamed by specialty at the outset with selection to all specialties being solely by performance in the structured *curriculum vitae* (CV), structured referee reports (RR) and structured, multi-station interviews (Int). The annual, single point of entry made selection to SET a very 'high stakes' assessment. A key impetus for the current study was a lack of objective confirmation regarding the effectiveness of these selection instruments to discriminate between applicants or to predict performance during training.

Environmental influences—principles of natural justice

Concurrent with the introduction of SET in 2007, RACS was occupied in addressing Australian Medical Council (AMC) accreditation requirements that underpinned its authorisation to deliver surgical training in Au and NZ. The AMC accreditation standards emphasised procedural justice in the form of open access, formal, 'transparent' procedures, impartial decisions, reliable information, rigour and fairness of selection processes, merit-based selection, and accountability (Australian Medical Council (AMC), 2007). The Medical Council of New Zealand (MCNZ) recognised the AMC accreditation standards, reinforcing the collaborative approach to medical training between Au and NZ.

The confluence of the introduction of SET and the AMC requirements influenced RACS' choice of selection and assessment processes and instruments. The specialty surgical training boards agreed to implement a single, staged selection process, with common items—structured CV, RR, and Int—implemented at the discretion of each specialty, within specified constraints. This approach gave the specialties flexibility to tailor selection processes to their requirements while increasing procedural transparency and adhering to accreditation obligations (Collins et al., 2007; Collins et al., 2008). It resulted in the content and 'weighting' of selection instruments varying between specialties and, longitudinally, each year within specialties. The choice and combination of selection instruments was influenced by *The Brennan Report* (Commonwealth Department of Health and Family Services, 1998), however, the validity and reliability of the selection instruments had not been verified in the local context. In addition, changes based on speculation rather than evidence limited compliance with principles of natural justice.

Governance and implementation of change

Governance by the RACS' Board of SET (BSET) enabled the Board in General Surgery (BiGS) to gain RACS' sanction to develop and customise selection practices within agreed parameters. A culture of review and modification resulted in annual adjustments to selection instruments and methods. Changes included modifying the scored elements and the scoring structures of the selection items, and reviewing the proportions that the CV, RR and Int comprised of the total selection scores. Such changes, apparent in the selection regulations and selection instruments reviewed in the current study, were presumably intended to maximise the efficiency and effectiveness of the process to discern those attributes considered to be the most important. However, changes were made to selection items based on conjecture, and the frequency and number of changes made it difficult to determine their effects. Impacts of changes were not formally appraised by BSET.

Generic selection parameters pertaining to all surgical specialties, and reviewed annually by the BSET, may have moderated impetuous changes to selection instruments or processes, but may also have reduced the capacity of any single specialty to respond to specific concerns. Proposed changes to selection instruments were presented for approval annually, and modifications to selection items occurred in all years under review in the current study ("Minutes of a meeting of the Board of Surgical Education and Training held at RACS on Friday 8th October 2010", n.d.). Trainees' performance during SET was monitored via examinations, work-based assessments and anecdotal reports. Examination content, standard-setting and implementation were tightly regulated and closely scrutinised by RACS' boards and committees; however, the delivery and monitoring of work-based assessments was less rigorous and less stringent.

Selection competencies and performance domains

The current study, as the first empirical measure of the impact of changes to RACS' GS selection processes, has identified and examined associations between selection and assessment performance outcomes in the surgical specialty of GS. This study has found mixed results regarding the predictive capacity of performance in selection instruments. Appraising other facets of selection and assessment—such as their educational objectives, their roles in RACS' training, their implementation and the impacts of people and environmental influences—adds perspective and context to the findings and has informed interpretation of the results and implications for practice. I will consider performance domains of RACS selection and assessment items and some influences—particularly principles of natural justice and person–environment fit—on the implementation of these items to contextualise these aspects within GS SET program.

Components of the selection and assessment items were classified by the experience, or RACS competency—collaboration and teamwork; communication; health advocacy; judgement and clinical decision making; management and leadership; medical expertise; professionalism; scholarship and teaching; or technical expertise (Watters and Civil, 2011; "Competencies | Royal Australasian College of Surgeons", 2017)—that they represented. The findings of this study reveal that CV components were least aligned and RR components were most aligned with assessments undertaken during training. The components scored in the CV were mostly directed to activities relating to the RACS scholarship and teaching competency. The RRs, broader in focus, were oriented to the RACS competencies of judgement – clinical decision-making, communication, collaboration and professionalism.

Educational objectives: Positioning RACS' selection and assessment items in educational theory—Bloom's taxonomy.

No selection or assessment item was explicitly differentiated into educational objectives, but applying Bloom's taxonomy (1956) retrospectively to the selection and assessment items may help to clarify the intended performance outcomes. This could reinforce or challenge associations between selection and assessment items indicated by alignments of RACS competencies, or that have been revealed by the statistical analyses undertaken in this study.

Selection items.

In RACS GS selection, achievements were assessed through the CV, abilities were assessed by the RR and Int. Candidates were ranked and admitted to training, with the number of trainees determined by the number of funded training positions available.

Structured curriculum vitae (CV). The paucity of positive predictive relationships identified between the Curriculum Vitae (CV) and assessments implies that there was little affiliation between components scored in the CV and those assessed in training. As indicated above, the CV primarily scored activities related to 'scholarship and teaching', a competency that formed only a minor component of the end of term assessments (ETA) and no component of the direct observation of procedural skills (DOPS), MiniCEX or the clinical examination (CE). Competencies assessed in the generic surgical science examination (GenSSE) and specialty surgical science examination (SpecSSE) were not specified, but it may be deduced that cognitive ability, consistent with the scholarship and teaching competency, was required by trainees in order to learn and retain knowledge about anatomy, pathology and physiology. Similarly, it is likely that the intention of the scholarship and teaching aspects of the CV were to attest to an academic, or cognitive ability in candidates that would fit them to undertake study to meet cognitive standards required in surgical training. However, the CV did not identify levels of expertise in scored experiences and CV scores simply represented records of candidates having completed certain activities rather than measuring the standards at which they performed these activities. The relevance to surgical training of activities scored in the CV—such as performance at elite levels in sporting or community service achievements—was not specified, but it may be assumed that these activities were considered to require attributes that would transfer to surgical training.

Speculating on why scored components such as courses, qualifications, clinical experiences and community and sporting achievements were valued, and considering the CV in the context of Bloom's (1956) three domains—cognitive (knowledge-based), psychomotor (skills-based) and affective (attitudinal based)—helps us to infer the possible intended educational counterparts of CV items. Courses in which participants undertook technical activities had a clear focus on psychomotor skills, while higher degrees were more likely to reflect the cognitive domain; clinical experiences could combine aspects of both these domains and sporting or community activities conceivably combined the psychomotor and affective domains. The CV scoring framework did not differentiate activities requiring more complex levels of knowledge, skill or attributes by allocating higher scores or a greater weighting; it is therefore difficult to situate the CV at particular levels within Bloom's domains.

High-scoring CVs represented collections of experiences, with little indication of performance standards or candidates' ability. As these experiences required days, weeks or even years to complete, candidates with high scores in the CVs had spent time and effort amassing qualifications, research publications and community achievements. In the competitive selection environment, using information from published CV scoring rubrics, candidates may have identified that they could directly influence their CV scores. However, this study revealed that attributes required to gain qualifications, conduct research or undertake community service were not directly assessed in the first two years of surgical training, and the experiences represented in high-scoring CVs had no clear alignment with surgical training requirements.

The structured CV contributed to the accountability of appointment processes—allocation of scores was clear and the methodology was publicly available. The CV attested to selectors and others that applicants had undertaken activities that denoted cognitive ability at a level that was likely to suit them to surgical training. For the CVs under review in this study, the scored activities and possible scores were clearly stated in selection regulations; however, we have seen that educational objectives in the CV were unclear and performance standards were neither stated nor scored. Furthermore, the scored items in the CV did not represent or align with activities in surgical training—a possible source of misinformation and frustration for candidates.

Structured referee report (RR). The structured referee report (RR), as the selection item most closely aligned to assessments during training, specified similar assessment criteria and performance statements to those in the ETAs. Predictive relationships between RR and ETA performance outcomes are supported by reviewing their stated competencies and identifying their educational objectives. The similarities of formats, assessment methods, question content, competencies and assessment measures in these reports reflect a high degree of association between the assessments. Categorising the reports' content according to Bloom's taxonomy (1956) is similarly straightforward with most content for both assessments pertaining to the cognitive and affective domains. These reports were much more nuanced than were the CVs—the range of qualities assessed was broader and behaviour that met or exceeded performance standards was rewarded with higher scores. Furthermore, RR scores usually represented the aggregate of five assessments, and candidates' behaviour was assessed in clinical contexts that bore similarities to those encountered during surgical training. Ideally, referees, in their roles as supervisors, had opportunities to observe candidates' behaviour over several months. This study was unable to ascertain the extent to which this presumed familiarity with candidates was the case, but it is reasonable to conclude that RRs were presumed to result from longstanding professional clinical relationships and that referees could identify attributes in candidates that were likely to suit them to surgical training. However, referees were not necessarily trained as assessors and although the selection methods minimised opportunities for partiality, the reliability of referees' assessments was not tested. The increase in RR mean scores over the three years under review may indicate an 'inflationary' scoring pattern that would decrease the reliability of this instrument. Mitigating this, the referee reporting processes provided multiple assessments of candidates' clinical performance, and candidates gained experiences during clinical rotations that were likely to give them insight into conditions that they might encounter in surgical training, increasing the likelihood of better person–environment cultural fit.

The RR, having most clearly defined assessment criteria and performance standards and representing assessments by multiple assessors of multiple behaviours, elicited over time, had perhaps the most face validity and potential for reliability of the selection items. This was borne out in alignments with the ETAs. As with the structured CV, this selection item reinforced selection process accountability—all candidates were assessed against the same, specified criteria. The nature of these work-based assessments ensured that candidates were exposed to clinical environments and assessments similar to those likely to be encountered during training, enhancing the person–environment fit of trainees.

Structured multi-station mini-interview (Int). The change from long-form semi-structured interviews to structured multi-station mini-interviews (Int) in 2009 resulted in increased associations between performance in the Int and in assessments for the 2009 and 2010 cohorts. The introductory open-ended questions in the 2008 interview were scored to a 'criterion statement' that required 'potential trainees to have a genuine interest and knowledge of General Surgery and to have an accurate perception of their own qualities'. While this statement is laudable it was not demonstrable, nor testable by questions such as "Why have you chosen General Surgery as your future career?" Subsequent questions in the 2008 interview showed similar lack of clarity about what was being tested and unclear connections between the criterion statements, questions and rating scales.

The 2009 Int tested five of the RACS' competencies: scholarship and teaching; communication and collaboration; management and leadership; health advocacy; and professionalism as well as "contribution to General Surgery". Questions and criteria statements were more structured and more closely directed to specific attributes than they were in 2008. Like the RRs, the 2009 and 2010 Ints addressed the affective and cognitive domains, although attributes and scoring standards were not as clearly defined as those in the RRs. The Ints had fewer questions than did the RRs, and were therefore less stringent and nuanced assessments that were the RRs. Furthermore, the RR scores were derived from more referee 'samples' of assessment, based on observations of performance over time, potentially increasing their reliability. Performance differences between the two selection items are evident when comparing numbers of predictive relationships—the RRs being aligned with more than twice as many assessments as the Ints.

Others have found that multi-station mini-interviews (MMIs) predicted performance in Objective Structured Clinical Examinations (OSCEs) Eva et al (2009) and Roberts et al (2008), in their studies of selection to medical school reported that MMIs of eight (Roberts et al 2008) to ten (Eva et al., 2009) stations predicted OSCE performance in Sydney and Ottawa. The results from the current study found a broader set of associations. Consistent with Eva et al's (2009) and Roberts et al's (2008) findings, the current study indicated that the 2009 (Au and ANZ) Int and the combined years (Au and ANZ) Int predicted performance in the OSCE format CE (2009 cohort and combined years). The current study additionally revealed that the Int was aligned with performance in the GenSSE (2010 and combined years; NZ), the SpecSSE (2009, ANZ; 2010, NZ; and combined years, NZ) and the ETAs (2008, NZ; 2009, NZ and ANZ; and comb. years, NZ) and was inversely associated with performance in one direct observation of procedural skills assessment (DOPS) (2010, Au DOPS3). See Appendix A Table 73. Cognitive components in the scholarship and teaching section of the Int may have enhanced associations with the SSEs, as both tested the cognitive domain. The negative association between the 2010 Au Int and DOPS3 may be an anomaly; no other association was identified between the Int and DOPS.

Embracing transparency of processes and natural justice principles, GS made the Int scoring processes publicly available, although the questions were not revealed to candidates until they were encountered at the interview stations. This common practice was likely intended to reduce candidates' ability to prepare answers to questions—minimising responses where candidates appeared to have prepared and rehearsed responses, as reported in studies such as Griffin et al., (2012) and O'Brien et al., (2011). Candidates' candid responses were sought in preference to constructed statements, tailored to what candidates assumed interviewers wanted to hear and were likely to score highly.

A risk to the validity of interview questions can be encountered when untrained selectors devise questions to elicit personality traits, or psychometric attributes. Studies have shown that when interview questions are based on unproven assumptions, the accuracy of the results is more likely to be questionable and unprovable. It can be enticing for selectors to conflate the outcomes of psychological tests with interviews but these are unlikely to withstand scrutiny or to comply with principles of natural justice.

Scoring of interviews was probably more subjective than scoring the CV or the RRs. Interviews scored a ‘performance’ in which candidates and interviewers participated, they were scored in ‘real time’ as they occurred or immediately after, and the GS Int questions and criterion statements varied in their precision and face validity. These factors combined to place more onus on assessors’ rapid judgements than did either of the other selection items. The written CVs and RRs could be scored and reviewed before being submitted and the RRs had more items and clearer performance statements, likely enhancing their validity compared to the Ints.

Assessing candidates’ unverifiable self-reports of past behaviour or projections of future behaviour may provide candidates with opportunities for ‘impression management’ (Culbertson et al., 2017; Ellis et al., 2002; Eva & Macala, 2014), whereby candidates say what they anticipate assessors would score favourably. Such actions compromise the reliability and validity of the interview. This has been addressed in some OSCE format MMIs and SJTs—approaches that can also expedite person–environment fit, by presenting candidates with scenarios that are concordant with surgical training.

Total selection score (Total sel). The total selection score (Total sel), being comprised of the CV, RR and Int, had elements of all three of Bloom’s (1956) domains. As the RR and Int comprised 80% of the Total sel, the affective and cognitive domains were most strongly represented in the Total sel. Unsurprisingly, associations between the Total sel score and assessments were most similar to those identified for the RR and were least similar to those observed for the CV.

Figure 35 presents a model indicating affinities between selection instruments and Bloom’s (1956) education domains.

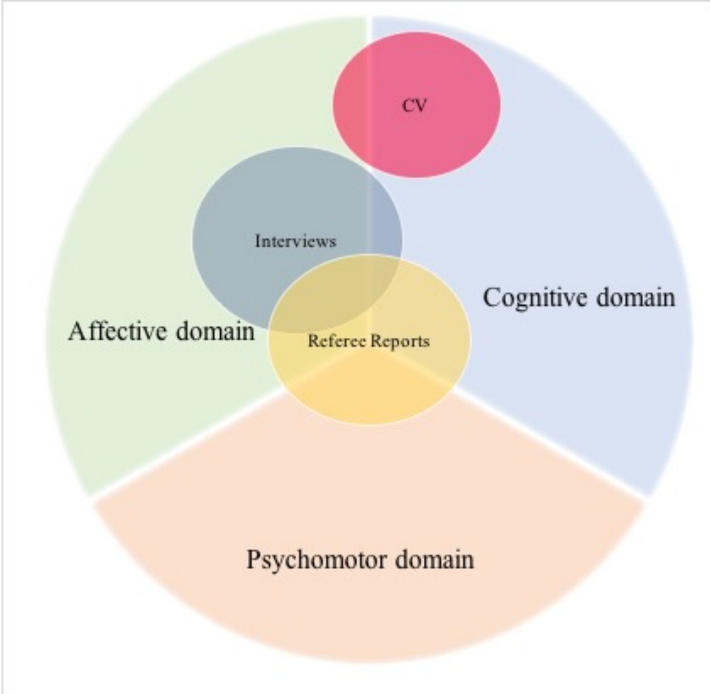

Figure 35. SET selection instruments and Bloom's education domains

Assessment items.

As noted for the selection items, the educational objectives of assessments were not identified by RACS but may be deduced from the content of the assessments.

Examinations. The generic surgical science examination (GenSSE) and specialty surgical science examination (SpecSSE), as tests of knowledge, reflected the cognitive domain. Questions within the examinations varied from low level knowledge recall to higher level analysis and synthesis. It is not surprising that few correlations were identified between selection items and these two examinations as the selection items chiefly scored performance in the two domains that were not represented in these examinations.

The CE tested the cognitive and affective domains in the four 'History taking' stations and the four 'Communication' stations, with elements from these domains also scored in the eight performance-based (psychomotor) stations. This echoes the composition of the RR—an observation supported by associations identified for the 2009 (Au) cohort—however, although this association was sustained in the combined years, no other Au cohorts, nor any NZ cohorts demonstrated associations between the RR and CE, so these associations cannot be seen as conclusive. The Int was more repeatedly associated with the CE than was the RR—possibly reflecting the emphasis in both the Int and the CE on communication skills. The 2009 Au (and ANZ) and the combined years' Total sel, predominantly comprised of the RR and Int, was aligned with CE performance, however no NZ cohort's Total sel was aligned with CE performance (see Results chapter Table 17). As trainees could undertake the CE in either country, at a time of their choosing, associative differences between the countries were more likely to arise from differences in implementing selection items than from differences in implementing the CE.

Direct observation of procedural skills (DOPS). As assessments of work-based skills, the Direct observation of procedural skills (DOPS) predominantly reflected the psychomotor domain—using sensory cues, implementing sequenced actions and demonstrating technical proficiency—but also drew on elements of the cognitive and affective domains in remembering, comprehending, analysing, evaluating, responding to phenomena, problem-solving and organising information. This might imply synergies with the Int, however, no positive predictive relationships were identified between the Int and DOPS assessments. Overall, very few predictive relationships were identified for the DOPS (see Results chapter Table 18).

Mini Clinical Evaluation Exercise (MiniCEX). The Mini clinical evaluation exercise (MiniCEX), assessing communication and professionalism, highlighted the affective domain, the patient examination component also reflected the psychomotor domain, while symptom analysis, diagnosis and patient management involved the cognitive domain. Performance at higher levels in these domains resulted in higher scores in the MiniCEX. Although very few predictive relationships were identified for the MiniCEX, the RR—the only selection item positively associated with the MiniCEX—similarly emphasised the affective domain.

End of term assessment (ETA). The end of term assessments (ETA), like the RRs, primarily assessed the affective and cognitive domains and were highly structured, identifying assessment criteria and providing performance statements to define standards; ETAs also grouped questions by RACS competencies, but included more competencies than were in the RRs. The affinity between these assessments was confirmed in the analysis. See Results chapter Table 20. See Methods chapter Table 10 for a comparison of the competencies assessed in the RRs and the ETAs. See Appendix D for an example of the RR form and see Appendix E for an example of the ETA forms.

Figure 36 presents a model indicating affinities between assessments and Bloom's (1956) education domains.

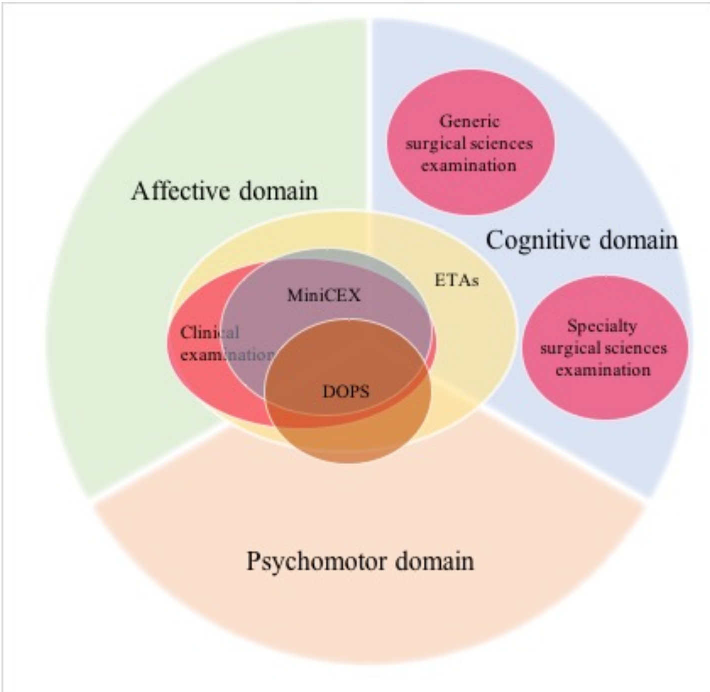

Figure 36. SET assessments and Bloom's education domains

Comparing selection competencies and performance domains with assessment competencies and performance domains.

Clarifying the educational objectives of the selection and assessment items highlights the nature of interrelationships that were identified in the statistical analysis. RACS assessments encompassed the multidimensional aspects of surgical training, readily according with all three (cognitive, psychomotor, and affective) domains described by Bloom (1956). Links discerned between selection and assessment items supported the notion of continued developmental progression by trainees through stages of increasing complexity in each of the domains.

The RACS' SET program was, and continues to be unique; the single point of entry made selection a very 'high stakes' assessment. Governance structures regulated procedural parameters, within which, changes to selection instruments and processes were the norm. These changes, proposed and effected by specialty training boards, were not based on empirical evidence. The validity, reliability, feasibility and fairness of selection assessments were unknown. Selection items comprised three complementary activities, spanning a limited range of knowledge and skills, while the suite of assessment tasks comprised many complementary activities, spanning a broad range of knowledge and skills. The validity and reliability of some, but not all, in-training assessments were established by RACS.

Part 3 Key individuals in selection, and assessment

There are people, central to all education systems; people devise, implement and participate in selection and assessment processes. People, with their vaguaries, and inconsistent and unpredictable behaviour, make selection outcomes uncertain. Three groups—the candidates, GS and RACS representatives and independent referees—interacted to undertake selection processes for GS SET. During SET training and assessment 'candidates' became 'trainees', independent referees were no longer required, GS and RACS representatives had administrative and/or assessor roles and employer representatives became involved. Governing bodies, including the AMC, MCNZ and RACS, set parameters delineating many selection and assessment activities. All the individuals—candidates, referees, interviewers, trainees and assessors—brought different experiences, skills and aspirations to surgical selection and assessments.

SET candidates declared themselves as aspirant surgeons, submitted themselves to the selection processes and were ostensibly prepared to commit to the GS SET regime; they had already undertaken several years training to qualify as doctors. GS and RACS representatives—usually surgeons and administrators—undertook to set, monitor and comply with selection, training and assessment regulations and to make judgements accordingly; referees—surgeons and other medical professionals—agreed (albeit tacitly) to honestly report their evaluations of candidates.

Balancing the competing tensions between the information required and permitted by GS, the efficacy of the instruments used, and the information that candidates sought to share about themselves was fundamental to GS selection processes. Candidates strove to present and promote their achievements and abilities in order to score—and rank—as highly as possible however, they were constrained regarding the information they could impart about themselves. GS representatives sought to gain accurate perceptions of each candidate's knowledge, skills and attitudes to establish the extent to which they met the selection requirements that were intended to represent the knowledge, skills and attitudes that would be beneficial in surgical training and practice.

Identification of desired attributes of trainees

Researchers have grappled with defining desired attributes of surgical trainees and with using these to inform selection criteria. Cognitive factors, physical skills (psychomotor) and personality traits (affective) all contribute to trainee performance (Dirschl et al., 2002). The specifics of attributes sought and testing methods vary per program.

If facilitating and producing proficient surgeons are major goals of surgical training, then basing training objectives on surgeons' optimal attributes would seem logical. However, we have seen that there is no 'one size fits all' description of a surgeon. Defining surgeons' attributes soon becomes a checklist, so long as to be almost meaningless. Furthermore, such a list of attributes is not static—it changes in response to societal and medical standards, attitudes and practices. The relatively 'high level', overarching nature of the nine competencies adopted by RACS (Watters & Civil, 2011) appears to be specific enough to convey meaning without being so detailed as to become trivial. However, the clarity and degree with which these competencies are integrated into the RACS GS training program are unclear. Some selection instruments and training assessments—particularly the RRs and the ETAs—are overtly based on the RACS competencies, but other assessments show little evidence of being so. This study did not review the manner or extent to which the competencies are taught.

Applied context of involvement in RACS' selection and assessment

The regulated selection practices implemented by RACS denoted that individual candidates, trainees, GS and RACS representatives and referees had defined roles. Constraining selection processes to three scored instruments with no latitude to include extemporaneous or supplementary materials reduced the influence of any individual beyond their defined area of responsibility. This is in contrast to the dominance once asserted by surgical 'masters' who directly chose their pupils and who were subject to influences of variable relevance to surgical training (Warren, 1951). As accountability and objectivity of policies, procedures and selection instruments have increased in importance, powerful and influential individuals have become correspondingly less important to the outcomes of selection.

GS candidates' actions were confined to providing specified information in the CV and in response to interview questions. Referees' influence was limited to assessing specified aspects of candidate performance. GS and RACS representatives' judgements were based solely on candidates' performance in the CV and interviews. Risks inherent in the selection processes related to the consistent implementation of selection instruments, the adequacy, relevance, accuracy and veracity of information, gathered about candidates and the capacity of assessors to use this information to make unbiassed judgements about candidates. From a candidate's perspective, there was a danger of being overlooked if selectors made incorrect, misguided or unwarranted inferences from the CV or interviews however, given the opportunity, some candidates might exaggerate or aggrandise their attributes and achievements to present themselves in their 'best' light and thereby gain an edge over their competitors (White, Brownell, Lemay & Lockyer, 2012). Candidates and judges alike placed trust in the selection instruments to gather apposite information. The candidates and the organisations trusted GS and RACS representatives and independent referees to make judgements about the candidates based on this information. This conferred great importance not only on the choice and implementation of selection instruments but on the impartiality, astuteness and capacity of GS and RACS representatives to discern relevant differences between candidates and to make judgements.

RACS' reliance on the CV and Int to convey personal information and on the RRs as assessments of clinical performance and/or as 'proxies' for personal recommendations reflected a desire to reduce the influence of individual, subjective judgements. However, as some researchers (e.g., Cone et al., 2012; Crossley, Davies, Humphris & Jolly, 2002; Norman, 2005; and Reiter & McConnell, 2014) have noted, many seemingly objective assessment measures include subjective components, particularly when they involve making judgements. Gathering data from a variety of sources—by including multiple selection or assessment instruments involving numerous assessments of performance by several assessors—decreased any individual assessor's influence on final scores, while increasing the validity of selection (or assessment) scores (Vassiliou & Feldman, 2011). The GS requirements for five RRs and for multi-station Ints in selection and for regular work-based assessments during training may therefore have contributed to the reliability of judgements and to the congruity of performance between the RRs and work-based assessments.

People are central to selection processes however, over time, the relative importance of individuals has decreased in selection to surgical training, as the importance of objective instruments and controlled procedures has increased. For RACS' GS this has resulted in greater standardisation of selection protocols with individuals performing defined roles (GS regulations 2008, 2009, 2010).

Procedural justice

Inherent in procedural justice principles are notions that increased fairness of processes will lead to increased objectivity and reliability of judgements. This study has confirmed that although the term was rarely used explicitly, compliance with procedural justice principles was fundamental to the evolving GS selection and assessment processes; *The Brennan Report* (Commonwealth Department of Health and Family Services, 1998), AMC accreditation standards (*AMC - Accreditation and recognition, 2017*) and worldwide trends (Carroll et al, 2009; Dowell, Lynch, Till, Kumwenda, & Husbands, 2012; Tsouroufli & Malcolm, 2014) underscored the RACS and GS emphasis on transparency, accountability and fairness in processes and decision-making. At an organisational level, GS selection processes conformed to commonly understood principles of procedural justice (Australian Health Practitioner Regulation Agency - Legal Practice Notes - LPN 17, 2013)—judgements were to be based on evidence and candidates expected a fair hearing from reasonable, unbiased assessors. Given the stated selection processes, any areas in which RACS or GS practice may have fallen short of procedural justice ideals were likely to have been unintentional—for example if the organisations (RACS or BiGS) lacked knowledge about best practice, or through individuals' limited or inconsistent implementation of the processes. It is possible that some candidates, referees or interviewers sought to manipulate selection outcomes by acting outside the system—to exaggerate or score individuals' performance inaccurately, to bypass the scoring mechanisms or by other measures—however, the selection processes provided limited scope for such activities to influence selection outcomes.

Although individual influences on selection outcomes diminished as reliance on procedures increased, people were still key to GS selection processes. Whenever subjective judgements were made, the potential to compromise procedural justice increased; individuals made judgements as they compiled content for the selection instruments and as they assessed candidates' performance. Appraising participants' adherence to principles of natural justice could identify compliance and shortfalls and provide a basis for future training.

Creating the scored items and establishing the weightings of these items within the selection instruments could reinforce biases or discriminate for or against certain population groups (Cleland et al., 2012; Yeates et al., 2012). As no such intention was identified in information published about GS selection, any discriminatory biases were likely to be pervasive and unacknowledged. This study did not evaluate the propensity for prejudice or bias in GS selection instruments. As the CV, RR and Int were authorised by GS' and RACS' committees, any biases in their content reflected institutional rather than individual prejudices; individuals could affect procedural justice when assessing and scoring candidate performance in the RRs and Ints.

The highly structured framework of the CVs presented little opportunity for subjective judgements and posed minimal risk to natural justice principles; the RRs and Ints, however, drew on many subjective judgements. Referees and interviewers might therefore benefit from training about compliance with natural justice principles. The RR**s** presented a challenge to assessor training as referees could be outside RACS' and GS' jurisdiction and therefore not subject to training requirements. Interviewer training has been promoted as a procedure to increase the reliability of outcomes (Salvatori, 2001) similar to assessor training increasing the validity and reliability of examinations and work-based assessments (Crossley et al, 2002). GS and RACS offered training to interviewers, but this study did not establish the extent or nature of this training.

Scrutinising and addressing any pervasive organisational biases may require regulatory, or organisational approaches, whereas individuals' biases might be addressed within GS at local or individual levels. The addition to selection regulations and to information provided to candidates of clear statements from RACS and GS that selection adheres to principles of natural justice could provide a focus for organisational reviews of selection questions, scoring and behaviour. Accountability is fundamental to procedural justice and fairness. RACS and GS ensured that selection and assessment policies and processes were scrutinised and authorised at the organisational level; external appraisal was conducted by the AMC through its college accreditation processes. Individuals' implementation of policies and procedures was less clearly defined and was not explicitly reviewed. This may have resulted in people inconsistently applying principles of procedural justice in selection and assessments.

Validity, reliability, feasibility, acceptability bias and fairness.

No selection instrument or process is inherently valid or reliable (Downing, 2003; Downing, 2004; Wiliam, 2001), rather, validity and reliability result from the ways instruments are implemented. Maximising the validity—particularly the construct validity and predictive validity—and the reliability of selection instruments contributes to procedural justice. This is particularly important in high stakes assessments, such as selection to surgical training (Downing, 2004). Predictive validity may be increased by using multiple tools in combination (Downing, 2003; Edwards et al., 2013; Patterson et al, 2009), by training assessors (Wiliam, 2001; Salvatori, 2001) and by using explicit standards and rating guidelines. Reliability may be increased by using clear, relevant test items and using multiple test items (Wiliam, 2001). Although reliability and validity may at times be in tension (Wiliam, 2001), there is general agreement that using multiple tests and carefully defining the constructs being tested increases the validity and reliability of selection or assessment outcomes (Dore et al., 2010; Edwards et al., 2013; Murphy et al., 2008; Williams et al., 2012).

One proposal to maximise reliability of selection is to introduce “more robust work-based assessment tools” to assess performance of candidates prior to selection (Collins, 2007, p. 9). Using results from structured work-based assessments in selection has the potential to enrich the information that candidates present with by adding authenticity. It may also enhance reliability by reducing the time lag between performance and assessment and may reduce pressure on assessors to complete and return reports to selection administrators. However, such assessment would be reliant on the cooperation of assessors, regardless of whether they had an affiliation to a specialty training program. To optimise this approach, assessment tasks and instruments would need to be predetermined so candidates could instigate assessments to build longitudinal portfolios of their achievements.

RACS GS used multiple selection items and considered the competencies tested by each instrument, however there was no evidence of interviewer training nor analysis of the reliability of selection items. The current study indicated that the RR and Int had limited predictive validity, and that the CV had inverse predictive validity.

Person–environment fit

Candidate admission into a surgical specialty training program is a reciprocal process. Candidates choose a surgical career from a range of career options, and surgical specialty representatives select trainees from doctors who apply. However, limited consideration was evident in the current study regarding the sources or quality of information used by ANZ doctors when they decided to undertake surgical training, or about the extent to which GS candidates reflected on their own suitability for surgical careers, or on the accuracy of their impressions of surgical training and practice. As described in several studies, (such as Bell et al., 2011; Go et al., 2012; and von Websky et al., 2012) if a candidate’s desire to train as a surgeon was based on inaccurate perceptions or information, or if their self-assessment of their skills and suitability for a surgical career were inaccurate, they could be at increased risk of struggling during training, potentially leading to failure in assessments, or of abandoning their surgical training or career.

Reasons for candidates' choice of specialty can be complex. Particularly in RACS' previous two-stage surgical training model, some aspirants may have chosen GS more as a vehicle by which to enter surgery than through a desire to be general surgeons. As GS had the largest annual intake, applicants had a greater chance of being selected into GS than into any other specialty. Some candidates may have appraised the expediency of transferring from the GS program if they were not initially selected into their preferred specialty. Others may have considered GS as a vantage point from which to observe features of other specialties, to which they might transfer should they develop a preference for an alternative specialty. In the interests of reducing attrition from GS, SET administrators discouraged this practice. SET regulations specified that trainees could not transfer between specialties, but were required to apply to be selected, competing with all other candidates. It was therefore in candidates' interests to gain adequate knowledge of specialty training programs and selection requirements prior to application and to apply to their preferred specialty at the outset.

The extent to which candidates knew of, or were able to identify with features of the training program or with the practice of GS, (the person–environment fit) was not measured in RACS' selection but it was likely assumed by GS that candidates were aware of both challenging and congenial aspects of the GS SET program. RACS GS selection practices do not appear to consistently evaluate candidates' prior knowledge of surgical training or practice however, adherence to Kelz et al's (2010) recommendation that information addressing "the realities of the surgical training environment" (p. 537) be made available would increase candidates' awareness of the GS approach to training and of the demands and rewards of a surgical career. Noted benefits of this approach include better management of trainee expectations, increased likelihood of better 'fit' between the person (trainee) and the environment (training program), increased trainee satisfaction and retention and decreased stress, burnout and attrition (von Websky et al., 2010; Schaufeli et al., 1996).

In the competitive selection process, consistent with recommendations regarding best practice recruitment and selection by Kelz et al (2010), and Klotz et al (2013), GS aspired to collect adequate, relevant and authentic information about candidates and to prioritise those whose behaviour and attributes were most likely to suit them to surgical training. Edwards et al (2008) identified that individuals who are most likely to meet required performance standards and be retained are those whose attributes most closely match the characteristics of a vocation. Reducing 'mismatches' in appointments to surgical training is predicated on the two-way flow of adequate, accurate, relevant information between candidates and selectors. RACS and GS provided information about selection, training and assessment processes on their websites, but did not formally review the extent to which trainees were prepared for or suited to surgical training. Inferences about candidates' 'fit' to surgical training were drawn from their performance in selection processes—including the implicit presumption that high scores in the CV and RR reflected experiences in situations that increased candidates' knowledge about surgical training and prepared them for this training. The more closely aligned the scored experiences in these selection items were to surgical practice and training, the more likely they were to accurately gauge candidates' exposure to and knowledge of surgical training. This was supported by correlations revealed in the current study between RRs used in selection, and the CE and ETAs conducted during training.

Interviews have been used to quiz candidates about their understanding of, and suitability for, medical training, however this approach has been criticised as encouraging prepared responses constructed by candidates to meet the perceived objectives of training, with little ability to verify the accuracy of candidates' statements (Griffin et al, 2012). Conversely, day–long 'work-experience-interviews', as described by Seabott et al., (2012), immersing potential surgical trainees in actual surgical environments, provide selectors and candidates alike with context–rich opportunities to judge candidates' fit with particular training programs. The feasibility of these full-day ‘interviews’, however, is limited, due to the intensive resources and time required; feasibility is particularly compromised when there are large numbers of applicants. Assessment centres, MMIs and SJTs straddle a middle ground, particularly when designed to test performance in key job-related simulations (Armstrong, 2006); candidates have commended SJTs for increasing their understanding of surgical training by presenting authentic scenarios (Koczwara et al., 2012) and candidate performance in assessment centres and SJT structured, behavioural-based interviews has been linked to later performance (Koczwara et al., 2012; Patterson et al., 2013; and Prager et al., 2010) implying that these assessments test attributes consistent with candidates’ fit to subsequent training.

The current study showed mixed results from the Int and did not explore candidates’ opinions of this methodology. However, it may be assumed that RACS and GS interviewers were satisfied with the MMI approach as it was retained through, and subsequent to, the years under review. The current study revealed potential for improving MMI implementation, particularly by monitoring and maintaining quality of questions and consistency between Au and NZ MMIs.

Candidates’, selectors’ and assessors’ performance

People’s roles in RACS selection and assessments were constrained by regulated processes. Candidates for selection and candidates for SET assessments were restricted in the communication avenues available to them. It is in candidates’ best interests to perform to the best of their ability to accurately represent their knowledge, skills and attitudes as they relate to activities being assessed. Individuals—whether candidates, selectors or assessors—who misinterpret instructions or questions, or who perform uncharacteristically are likely to affect the reliability of their assessment results. Any single selection referee’s or interviewer’s influence on selection outcomes was moderated by being one of several judges; however, the skill with which individuals undertook their roles had some potential to affect selection outcomes. The influence of individual assessors was greater for work-based assessments as, particularly for the DOPS and MiniCEX, lone assessors conducted these. Inconsistent implementation of policies and procedures or biased or unreliable judgements could also compromise selection and assessment practices and outcomes. As Downing (2004) reminds us, the quality and consistency of judgements made by 'human raters' affects the reliability of assessments, with the largest threat to the reproducibility of clinical assessments being rater inconsistency.

Performance standards for referees, interviewers and assessors were not specified in GS selection or assessments. However, training that addresses principles of natural justice, educational principles of assessment and the requirements of each selection and assessment role is logical to consider to maximize the reliability of selectors' and assessors' ratings. Training could identify standards of performance required of those entering surgical training, thus helping referees and interviewers to calibrate their judgements appropriately. An additional conclusion is for candidates to be thoroughly informed about selection and assessment goals and processes, as the more candidates understand the goals and the more familiar they are with the processes and what is required of themselves, the more likely they may be to respond to designated selection and assessment criteria rather than imparting irrelevant information or orienting themselves to the processes during selection and assessments.

Objective policies, procedures and measures; subjective judgements

Optimal selection outcomes balance objectivity and subjectivity. Policies, procedures and instruments are often designed to be as objective, impartial and fair as possible. Care must be exercised to ensure that this objectivity is not illusory. When variation in assessment content and experiences are minimised, it is assumed that differences in outcomes reflect differences in candidate performance. Structure and uniformity are central, however, moderating factors, such as affirmative action to redress inequity, although essentially subjective, can enhance procedural justice and fairness.

Individuals, conveying and interpreting information, and making judgements, generate the most subjective elements in selection processes. Candidates interpret what is required of them in the selection process, and strive to present themselves to establish an advantage. In high stakes and intensely competitive selection processes, such as the RACS' GS selection, candidates' judgements about what will place them ahead of others are not only subjective, but may encourage them to tailor their comments and behaviour to conform to what they believe is expected. Such responses may not provide an accurate portrayal of their normal behaviour. Assessors' judgements are influenced by their biases and inconsistencies. Assessor subjectivity may be mitigated through activities such as assessor training, by providing assessors with feedback on their performance, and by including multiple assessments, with multiple assessors in selection processes.

Some performance measures that are ostensibly 'objective' may discriminate between candidates on grounds other than those intended (Cleland et al., 2012). This form of discrimination is particularly likely to affect candidates who come from disadvantaged backgrounds, or who have had limited exposure to preparatory activities or experiences. Candidates do not come from an 'even playing field'. It is important to consider that selection instruments are as fair as possible. Fairness is more than the extent to which an assessment follows principles of procedural justice. Fairness encompasses and attempts to balance merit, access, diversity, processes and outcomes, judged against a background of social ideologies and expectations.

Candidates' perceived merit for selection into surgical training is measured and assessed through their performance in selection instruments, but may also take account of social, cultural, racial or other attributes. Compensation for disadvantage or inequity of access to preparatory experiences may be applied in addition to performance in selection instruments (Cleland et al., 2012; Gipps and Stobart, 2009). Judgements regarding the intent and implementation of compensatory measures must be communicated to all participants. Review of the effects and outcomes of compensatory measures will feed into the cycle of selection development and implementation activities. Selection to surgical training benefits from the inclusion of both objective and subjective measures to assess candidates.

A major challenge is to implement processes that are truly 'fair' and to behave fairly to all. Candidates expect that selection processes and allocation of training positions will be conducted fairly, according to principles of procedural justice, and that outcomes will be merit based (Razack et al., 2014). Unhooking 'merit' from 'academic achievement' in competitive selection opens selection processes to a multiplicity of influences and tensions—between fairness to candidates, to society, to the surgical profession and to specialty training boards. Judgements about the relative fairness of activities are multifaceted; maximising fairness in selection to surgical training requires adherence to principles of procedural justice and consideration of values, protocols, procedures, performance standards, assessment methods, tools and possible outcomes. Such judgements may "seek excellence … in equitable and diverse ways" (Razack et al., 2014, p. 43). Additional complexities in judging the relative fairness of selection activities arise when measures to redress disadvantage result in the admission of trainees who are poorly equipped to complete surgical training. Subsequent supports for these trainees may be warranted.

Clear public statements regarding the intentions of selection protocols and instruments will strengthen transparency and the procedural justice aspects of selection. Providing comprehensive information to familiarise all candidates with the selection format and activities would also contribute to equity of access by facilitating and directing candidates' preparation activities. Similarly, organisationally-endorsed coaching that is broadly available could contribute to the fairness of selection for all candidates. A caveat to this, however, is to ensure that all those who might benefit from such coaching can access it.

In current ANZ surgical training environments, fairness, merit and diversity are linked through notions of 'inclusion', which have become increasingly important to RACS during the years since this study started (*Royal Australasian College of Surgeons Accreditation Submission 2016*, 2016). One example of this is the development of the RACS Diversity and Inclusion plan (*Royal Australasian College of Surgeons Diversity and Inclusion Plan,* 2017). Safe and inclusive selection practices seek to address historic inequities in the composition of the surgical profession, and in the notion of 'merit-based' selection. Implementation of inclusive selection practices will better represent the diversity of ANZ society and are likely to encourage diverse voices to speak, be heard, and to challenge entrenched views (Razack et al., 2014). There is currently a tension in the discourse between the relative importance of diversity and selection based on 'merit'.

Part 4—Utilising selection instruments and protocols

Evidence from others' research and from the current study supports the use of multiple instruments in combination, to differentiate between candidates in order to admit those appraised as most suitable for surgical training. However, the optimal number and mix of selection instruments and implementation processes is not clear-cut; there is no definitive combination. Preferred options will be context–specific, dependent on resources and personnel available, on local precedents, on awareness of alternatives and preparedness to critically consider and implement them, and on the intended purpose and required outcomes of selection. Resolving these considerations into a "cohesive and comprehensive model" that is grounded in theory and is legally and ethically defensible is challenging (Bore et al., 2009, p. 1066).

Academic achievement, the erstwhile cornerstone of selection practices, has been shown to be a limited predictor of performance in specialty medical training. Tests of academic ability or knowledge are unlikely to permit meaningful judgements to be made to differentiate between closely-matched, high-performing candidates, such as those who apply to surgical training. Furthermore, such tests may discriminate against candidates who come from disadvantaged backgrounds or who have had limited exposure or access to beneficial preparatory experiences or resources. Although RACS GS did not include a test of academic ability in the selection processes reviewed in the current study, the CV included points for educational and research accomplishments, an aspect that bears comparison with academic achievements. The current study revealed the CV as the selection instrument least aligned to performance in surgical assessments.

Attributes other than academic achievement have been deemed important for success in surgical training and practice. Typically, these may encompass non-cognitive constructs, personality traits, visuospatial, and psychomotor skills. There is broad, general agreement that attributes such as conscientiousness, judgement, dexterity and communication skills will benefit surgeons and trainees, however, absolute taxonomies of non-cognitive attributes and their relative importance are likely to be limiting and illusory. Many analysts and authorities concur that an ideal medical workforce reflects the society in which its members practice. In Au and NZ, changing social mores and ongoing immigration have resulted in increasingly socially mobile and ethnically diverse populations. In addition, the practice of surgery itself is becoming increasingly complex, as new areas of expertise emerge, new procedures and approaches are developed, and new priorities are identified. This indicates that characteristics suited to the medical workforce are not fixed, and that a variety of attributes will prove beneficial for surgical practitioners (Bann & Darzi, 2005; Wanzel et al, 2002)). Attributes such as flexibility and adaptability are likely to assist surgeons to meet changing demands.

Selection activities may be able to assess candidates' current performance, but their capacity to assess candidates' latent ability or to directly 'predict' future performance is questionable. Studies have been undertaken to review students' capacity to learn cognitive and psychomotor skills and professional behaviours (Arnold, 2002; Baldwin, 2003; Buckley et al., 2014; Cruess et al., 2009; Kulatunga-Moruzi & Norman, 2002; Tansley et al., 2007). Although there remains disagreement among researchers, there is evidence that all these aspects can be learnt and developed, particularly by using 'deliberate practice' involving targetted training, feedback, repetition and 'effortful' concentration (Ericsson, 2015). Visuospatial ability appears to be more crucial initially and its importance decreases as practitioners gain experience performing a procedure (Tang et al., 2014). Practice and experience appear to obviate the impact of 'innate abilities' (Wanzel et al., 2003). This places the onus on those responsible for training to employ training strategies to facilitate trainees' skill acquisition. If identifying candidates' innate abilities at selection is de-emphasised, then selection panels may focus on defining and assessing minimum standards of proficiency in skills appropriate for trainees at the commencement of training.

The design and implementation of selection instruments also continue to evolve. Emerging practices, such as modifying assessment instruments—such as 'entrustable professional activities' (EPAs) (ten Cate, Tobin and Stokes, 2017)—for use in selection continue to enlarge the repertoire of instruments and techniques available to selectors. The use of EPAs in selection is yet unproven, but may allow assessments of authentic workplace activities to contribute to the repertoire of selection practices. Observations and discussions regarding the validity (face, criterion, construct, incremental, and predictive), the reliability, fairness, acceptability, and feasibility of myriad selection instruments singly and in combination have increasingly occupied medical educationalists and researchers. Internationally, research to provide evidence regarding the efficacy of the breadth of current selection practices continues to be undertaken (Cleland et al., 2012; Gallagher et al., 2014; Maan et al., 2012; Makdisi et al., 2011; Prideaux et al., 2011; Roberts & Togno, 2011; Schaverien, 2016). The evidence does not support any test used in isolation as a quintessential instrument for selection to surgical training. It is likely that, rather than specific instruments being ‘better’ or ‘worse’ in selection of surgical trainees, the combinations and incremental groupings of instruments more greatly affect the efficacy of selection outcomes.

The current study has highlighted that the ability to evaluate the effectiveness of particular combinations of selection instruments is compromised if they are inconsistently implemented—over time or by location. To achieve, enact, and maintain meaningful and effective arrays of selection processes and instruments in changing environments it is incumbent on those responsible for selection to establish the aims, approaches, and outcomes of their selection activities and then to routinely reconsider them. Effective combinations of instruments will reflect locally identified needs and global good practice. Confirming these aspects will support regulators to evaluate the relevance and effectiveness of their own methods and to determine the suitability of other approaches. Establishing initial robust selection protocols that are tailored to the local environment is likely to moderate the need for frequent modifications, encouraging a consistent and predictable model for selection. To address uncertainties and ambiguities inherent in ongoing changes to selection, introduction of any proposed changes may be graduated by piloting new instruments and appraising their performance before they are implemented. This approach could also be seen to promote fairness, by being consistent for candidate cohorts, and would support subsequent scrutiny by allowing reviewers to more clearly see any trends in the performance of the selection instruments or processes.

Custom influences the choice of many selection instruments; however, longevity is no guarantee of efficacy. The use in selection of grade point average (GPA) from previous courses, for example, may have derived from the assumption of academic performance being a proxy for intelligence or ability. Emery and Bell (2009) however, identified that prior attainment of high-achieving candidates can be a poor discriminator, lacking variability when almost all applicants are closely matched. Although these criticisms could be equally valid in relation to using tests of knowledge to rank candidates to specialty training, there may be a place for knowledge tests as prerequisites—to establish that candidates have amassed necessary knowledge to support further training. Instead of using knowledge-based tests as surrogates for other attributes, they could be used to ascertain whether candidates can meet the required standards of factual knowledge in specified disciplines.

GPA is a crude measure of one of the desired attributes—cognitive ability—in a doctor's repertoire. Appropriate combinations of more sophisticated selection instruments are likely to give more nuanced and meaningful results over a broader range of attributes. However, this presents selectors with a conundrum. Identification of 'desired' attributes of doctors may make selection activities more targeted and efficient—by enabling refinements to selection processes to better reflect core attributes—but neither the attributes nor the best ways of recognising and testing for them are universally agreed (Cuschieri et al., 2001; Dirschl et al., 2002; Farkas et al., 2012; Grantcharov & Reznick, 2009). Although it is important to reflect societal expectations, defining key attributes of good doctors is susceptible to convention, or fashion, and to 'pseudo-scientific' approaches. These are likely result in compromised selection activities with low validity and reliability. There is also a risk of adversely limiting the candidate pool by ignoring some potentially beneficial attributes or by not prioritising attributes appropriately, just as once it was popular to prioritise cognitive ability over attributes such as communication.

In the transition from elementary to complex medical training and practice, the approaches to learning and practice change from acquisition of knowledge and basic skills to synthesis and application of knowledge and skills as learners and practitioners diagnose, make clinical decisions, plan, treat, manage and communicate with patients and health professionals. To be relevant at each stage of training, selection and assessment activities must reflect these changes in emphasis.

Summary to the Discussion

The current study has identified aspects of the selection process that were aligned with the training program and some which had little connection to training outcomes. The methods used in this study were consistent with similar studies conducted elsewhere. A pilot study of Paediatric Surgery training data confirmed that the methods were appropriate to extend to this full-scale study of General Surgery training data. The methods revealed relationships between selection and assessment items, as reported in the Results chapter. The methods could be extended to include results from subsequent examinations and results from trainees in other surgical specialties.

This study has found numerous significant relationships between trainees' selection scores and their subsequent work-based assessments and academic and clinical examination results undertaken during the first two years of a surgical training course. The individual predictive ability of any of the selection measures—CV, RR, Int and Total sel—appears weak to moderate, with most (54%) of the statistically significant correlations being in the range 0.16 to 0.29 and only a few correlations exceeding 0.50. However, in organisational psychology, it is recognised that even quite weak correlations of cognitive, non-cognitive and behaviour measures may be useful predictors of workplace outcomes (Adam et al., 2012).

The research findings contribute to RACS' ongoing evaluation of its selection and training practices by comparing the efficacy of the selection tools and processes used by each surgical specialty. The findings will also enable each specialty to review its selection practices for their alignment with assessments undertaken during training and, should misalignments be identified, to modify the selection methods to more closely resemble those techniques which are most strongly linked to the early assessments.

Delimitations and limitations

Delimitations

Some limitations were placed on the study design, before the project commenced. These primarily relate to the study population and the data sets. Data were obtained from one surgical specialty (GS), in the surgical education and training program practised in Au and NZ. The study was limited to three annual cohorts of trainees, with assessment data pertaining to the first two years of training only. The findings from this study, considering trainee performance in selection and early assessments, may be supported or challenged by comparisons with trainee performance in assessments undertaken later in training, or from trainees in other surgical training programs.

The study only reviewed trainees' first attempts at examinations, to provide a common basis from which to compare examination performance. Trainees who failed an examination could be permitted up to three further attempts (see Appendix B, Examination Policies) at that examination. Reviewing examination performance of trainees who had multiple attempts at examinations may reveal other pertinent performance information.

Additional delimitations arose from the confidential nature of some data. It is not possible to reproduce detailed interview or examination content in this study as this content is not in the public domain and these selection and assessment items may be re-used subsequent to this report. Articulating the content of these items may adversely affect their future validity, reliability and fairness. The original dataset cannot be deposited in a publicly available repository as it contains information relating to confidential assessment data. This data is retained in a confidential digital environment.

Limitations

There are several limitations to the data used in this study. Data were collected retrospectively and some data points were missing. No demographic data were available from RACS on trainees. Selection score data were collected in all selection items and assessment performance data were collected for all available assessments. Assessment performance data for trainees selected in 2009 and 2010 (commencing SET1 in 2010 and 2011 respectively) is incomplete because, at the time of data collection, these trainees had not yet undertaken assessments that were implemented in the second and third years of training (SET2 and SET3). Some assessment data points had small numbers of subjects. These are noted in the results. The DOPS, MiniCEX and ETAs were implemented as formative assessments, with letter grades and assessor comments. Converting letter grades to numeric scores for analysis in this study ignored assessor comments and may have thereby have lost some of these assessments' nuances.

Although data were collected for all trainees entering GS SET in 2008, 2009 and 2010, it is not known to what degree the trainees whose performance data was used in this study typified the broader population of RACS GS trainees (who were selected in other years) or surgical trainees training elsewhere. Originating from one surgical training program, the findings may not be generalisable to other surgical specialty training programs or to other surgical training organisations. The findings from these three cohorts may also not be generalisable to other GS cohorts, particularly as selection instruments and protocols were subject to annual changes. When reviewing and discussing the results of analyses, it must be noted that the low number of NZ trainees (n = 56), may generate less reliable results than those identified for Au, which had considerably more trainees (n = 291).

It is acknowledged that using a small sample population increases the chance of assuming a 'false' premise is true, reducing the confidence level of the study. The results of this study are not presented as conclusive. Assumptions could be further tested with subsequent cohorts of GS trainees, providing additional longitudinal data.

A potential limitation was restriction of range for the independent variables. Candidates with low selection scores were not admitted to surgical training and therefore it was not possible to compare their selection performance to assessment scores.

This study did not address qualitative aspects of selection, such as would be gained by exploring participants' motivations and reflections on their experiences, or the degree of compliance or 'agency' adopted or desired by participants. It is possible, for instance, that individual candidates, referees or interviewers sought to manipulate selection outcomes by acting outside the system.

Changes to selection content and processes were recorded by the BSET. Discussion points and rationale for changes were not recorded in BSET minutes, however, and thus were not available to the current study.

A possible limitation exists regarding the currency of the data as this study was undertaken part-time throughout the process. Since the study commenced, more recent data have become available and RACS' selection methods have continued to evolve. However, the three selection instruments, the work-based assessments and the examinations reviewed in the study are all currently in use in the RACS' surgical training program.

Implications and recommendations for future practice

Several recommendations for practice stem from the findings of this study. It is anticipated that these recommendations will be presented for consideration by the RACS' and specialty boards and committees that are responsible for selection to training. The leading recommendation is for RACS to employ a coordinated, systematic approach to selection, framed by explicit, guiding principles that define the organisation's priorities regarding selection. Considered design and implementation of selection and assessment activities is essential and will benefit from integrating diverse perspectives. RACS would follow its customary methods to identify and articulate selection principles to suit local conditions. As a basis for deliberations, this study has established that it would be beneficial for RACS to consider maintaining and enshrining activities that strengthen procedural justice in selection; to advocate valid, reliable, feasible, evidence-based selection activities; to continue the use of multiple complementary selection activities and instruments; to promote diversity among cohorts selected into training, and to foster consistency between selection and assessments throughout training.

To provide a focus for surgical selection, training and assessment activities, it is recommended that RACS define the goals of its surgical training programs and of its selection processes. It may be helpful to use a taxonomy such as Bloom's (1956) to categorise and articulate the educational objectives of activities and assessments undertaken in selection and training. Such a taxonomy could assist RACS to describe the constituent elements and performance standards of the knowledge, skills, and attributes required of trainees at entry and at stages throughout training. Recognising that assessment encompassing the cognitive, psychomotor and affective domains could assist selectors to match candidates' observed performance with desired trainee qualities, at appropriate levels could enhance selection processes and outcomes.

A recommendation for RACS to maximise the effectiveness and efficiency of selection into all surgical specialties, presupposes that RACS will use performance evidence to amplify activities and processes that align with identified surgical training outcomes and will modify or remove instruments and processes that have been shown to be mismatched to training. This study ascertained that performance in the CV was a poor predictor of performance during RACS' GS training, and that performance in the RR, Int and Total sel were more likely to be aligned with subsequent performance in the first two years of GS SET. The CV, via its relatively large SD, was found to discriminate between candidates better than did the RR or Int. It is therefore recommended that RACS GS review the content of the CV to score attributes that are valued in surgical training. RACS could also consider removing the CV as a scored selection item; some CV elements, deemed to represent essential pre-surgical training experiences could, instead, become pre-requisites for GS SET. This approach would enable alternative selection instruments to replace the scored CV. As the RR and Int discriminated poorly between candidates, it is recommended that RACS GS review the content and scoring of these selection instruments to build their discriminatory power. This could involve adjusting question content, scoring scales and/or training assessors.

Additional activities to support revised selection practices include setting unambiguous parameters for surgical training and selection by clarifying the objectives of the RACS surgical training programs; by identifying the knowledge, skills and attributes that are valued in surgical trainees and in candidates to surgical training; and by clarifying the intent of selection. Such activities would maximise the likelihood that selection was clearly aligned to training, assessments and the professed objectives of training.

Once selection parameters are specified, it is recommended that RACS reinforces quality control of the development and administration of selection activities. To optimise selection instruments and determine their compatibility with identified objectives, RACS could analyse its own and others' selection instruments and protocols. Incorporating work-based assessment instruments (such as EPAs) into selection could improve the validity and reliability of assessments and be cost and time efficient. Further study would be required to determine the extent to which this approach would assist selectors to differentiate and rank candidates. Additionally, RACS could review recently developed selection and assessment activities that may not yet be widely adopted, to ascertain their suitability for use in the ANZ surgical selection context. It is anticipated that such a comprehensive review would result in RACS endorsing selection instruments and procedures that represent verified good practice and align with the objectives of surgical training.

Maximising the performance of participants—candidates, referees, interviewers, and assessors—is likely to enhance the reliability of selection assessments. Training for selectors and assessors would highlight the objectives and requirements of selection and of individual roles within the process. Content could include principles of natural justice, factors that affect the predictive validity, reliability and fairness of assessments, and performance standards required of referees, interviewers, administrators, assessors and examiners.

Similarly, providing clear and comprehensive information to candidates about the profession of surgery, about the ethos and implementation of the training program, about selection and assessment objectives processes, rationale, and about participants' roles in selection may not only enhance candidates' knowledge of surgical training, and contribute to optimal person–environment fit, but may encourage applications from candidates who have characteristics that are compatible with SET (Kelz et al., 2010) and assist candidates to maximise their selection scores. The greater candidates' understanding of selection objectives and processes, the more likely they are to provide the information required. It is recognised that although provision of training and information to all participants is likely to enhance their performance, thereby increasing the reliability of assessments, such a comprehensive approach may not be feasible.

RACS and GS could consider reviewing the information they provide to better describe the demands and rewards of surgical training and the educational style and culture of GS SET. This could include making resources such as structured curricula available prior to training. Such a change of perspective would denote selection as an opportunity to exchange information, rather than to 'extract' information from candidates. In an effort to increase the capacity of selection to provide the "insight and understanding of what is required" that Burgess et al. (2014, p. 4) recommend and thereby maximise person–environment fit, RACS and GS could consider introducing selection items that reflect the surgical training environment and challenge candidates' knowledge of surgical training. Selection items to consider could include SJTs and EPAs.

It is recommended that, where possible, revised or new selection practices are tested in parallel with current practices to ascertain their performance in the local context before being formally implemented; then, notwithstanding the importance of regular review, that the new practices be permitted to become established with few or no changes to content, format, or protocols. Such an approach would enable participants to become accustomed to the suite of selection practices and would help to reveal performance trends when selection activities were evaluated.

A recommendation to measure candidates' performance against standards (criterion referenced), rather than against other candidates (norm referenced) would contribute to procedural justice by defining performance standards and could contribute to equity of access and increased diversity among trainees. A criterion referenced approach would specify performance standards for selection and would likely necessitate defining minimum requirements—including a 'cut score', below which no candidate would be offered a training place. Such an approach could assist RACS to promote diversity if it were to reserve a percentage of places for nominated categories of applicants who meet the minimum requirements for selection, but who may not rank as highly as other candidates.

It is also recommended that RACS continue to monitor and evaluate selection practices by regular quantitative and qualitative reviews. RACS could repeat and extend correlational analyses such as the current study to compare selection performance with performance in later work-based assessments and in the Fellowship Examination. In addition, RACS could use performance data to evaluate the reliability of selection assessments. Qualitative studies could gather information from participants in selection processes to ascertain their reactions and impressions regarding the acceptability, fairness and feasibility of selection instruments and processes. Cyclical, structured, evaluation and review would contribute to evidence-based refinement of RACS' selection practices.

The findings of the current study will be made available to RACS to contribute to the continuing refinement of selection practices. The findings from the current study may be cautiously generalised to other members of the RACS community, however there is no intention to extrapolate more broadly.

Recommendations for future research

As the first study of selection to ANZ surgical training, the current study has identified many questions that could be addressed in future research. To be of maximum benefit to RACS, data could be analysed to establish the current performance of RACS selection practices across all of RACS' surgical specialties. Such research, extending the methodology developed for the current study, could review the predictive validity of selection instruments and protocols currently used by all RACS' surgical specialties. Data are available for all surgical specialties, from all cohorts selected since 2008. Longitudinal studies could now compare performance in selection against performance in assessments undertaken throughout surgical training, including performance in the final major assessment, the Fellowship examination. Identification of similarities and differences between cohorts, between countries and between surgical specialties could be explored. Such studies could identify trends and would inform the generalisability of the findings within GS.

Further studies could explore the accuracy and extent of candidates' familiarity with and preparation for surgical training and whether this is associated with trainee performance in SET. Reviewing the content, consistency and scoring mechanisms of the GS Int scenarios and questions and investigating their relevance to subsequent training and their acceptability to interviewers and candidates would provide further insights into the selection process, potentially increasing the likelihood of maximizing the fit between GS trainees and SET.

The increase noted in RR mean scores may be evidence of a trend to 'score inflation' whereby referees give unwarrantedly high scores to enable their preferred candidates to rank higher than other candidates. Future studies could monitor RR scores to discern whether this trend has persisted.

The outcomes of SET GS training are assumed to be near identical for Au and NZ. As this study revealed differences between Au and NZ selection and assessment performance, future studies may consider the extent of similarities and differences in selection, training and assessment practices between the two countries. The outcomes of such studies could have implications for practice undertaken in each country. The appraisal of participants' adherence to principles of natural justice could also identify compliance and shortfalls and provide a basis for future training.

This study did not consider the history of applicants to surgical training, particularly with regard to prior (unsuccessful) applications to GS, or applications to other surgical specialties. Further studies, using data reflecting unsuccessful applications, could reveal whether trainees' performance during training reflected performance during prior selection attempts.

Importance

This is the first study to objectively evaluate relationships between selection instruments and the assessments during surgical training in Au and NZ. It is anticipated that by providing a clearer understanding of selection processes and insights into relationships between selection and early assessments in GS training in Au and NZ that the current study provides data to inform changes to selection and to support continuing use of selection instruments that demonstrate positive links to training outcomes.

All trainees' scores were calculated for each selection item and, where available, scores were recorded for sub-components within the selection items. Assessment scores were recorded for trainees' first attempts in each of the three examinations—Generic SSE, Spec Spec SSE, CE. DOPS, MiniCEX and ETA reports for all trainees for each rotation were reviewed and ratings were converted to numeric scores for each report; scores for sub-components within the ETA reports were also calculated. These instruments were chosen to provide the data for this study, as they comprise the key, quantifiable indicators that are used in the GS training program to measure trainee performance.

Conclusions

Admitting candidates who are unsuitable for surgical training may result in trainees who struggle or are unable to satisfy training requirements and who may ultimately jeopardise patient safety and outcomes. Not accepting candidates who are well-suited to surgical training may be particularly unfair to those candidates and to the community. It may be assumed that all candidates have invested considerable time, resources and emotions into their applications; when those who are most likely to succeed in training are not selected these efforts are wasted and society misses out on skillful surgeons.

Continuum of development

Surgical training forms part of a continuum of development for surgeons. Prior to undertaking surgical training, individuals engage in formal education and other experiences that contribute to their readiness to commence surgical training. Surgical training programs build individuals' knowledge of, and skills in, surgical practices and assess their performance in these practices. Completing a surgical training program marks the closure of a stage, at which point it is recognised that individuals are authorised to practise independently. Qualified surgeons continue to improve their efficacy as they extend and refine their knowledge and skills throughout their careers.

The notion that attributes of excellence in surgeons can be used to frame training and, by extension to identify selection requirements is attractive. This approach has served society tolerably well for many years. However, descriptions of surgical proficiency differ and the 'traditional' western understanding, weighted to medical knowledge and technical dexterity, now appears limited. Accomplished surgeons do draw on extensive knowledge of anatomy, pathology, physiology, investigative and diagnostic methods, and the causes, effects and treatments for disease and injury, as well as being skilled in fine motor dexterity and hand–eye coordination; however, previous measures of proficiency ignored personal characteristics and interpersonal attributes that are now considered integral to clinical excellence (Schaverien, 2016).

Broadening the parameters of clinical excellence raises conundrums in determining surgeons' and trainees' ideal attributes and in selecting, training and assessing in relation to these attributes. It is tempting to seek simple solutions or to ignore behavioural or other aspects that may increase the complexity of these questions and solutions; however, oversimplification raises the likelihood of inadequate, unsound explanations. The appeal of simplicity, for example may be inconsistent with complex performance requirements, or with selection, or assessment instruments' reliability. Ultimately, the optimal, practical solutions to selection, training and assessment will be the simplest, most feasible options that maintain acceptable validity and reliability in local contexts.

The continuum of surgical training is comprised of many discrete phases, events and activities, undertaken in varying locations by countless participants and subject to many influences. One way to appraise this continuum is to identify the component parts to describe and, if possible, measure their interrelationships, interactions and outcomes, to comment on their efficacy and to recommend preferred ways to optimise progress through the continuum. This study has engaged with selection to specialty training—one of the junctions between two phases in the continuum—to review the extent to which candidates' performance in this activity was linked to their performance in subsequent training activities. The study addressed other phases and activities that impinged on, influenced or were pertinent to selection. The study has revealed the effectiveness of selection processes in recognising and admitting those who are most suited to surgical training in Au and NZ.

Within the continuum of surgical training the evident ambiguity and unreliability of some components is at odds with a desire for quantifiable cause-and-effect explanations and outcomes. The many complex interactions in surgical training and practice do not readily lend themselves to a predictable, causal nexus. Human agency, competing influences, inadequate specification of requirements and imprecise measurement instruments and processes are among the most obvious contributors to inconsistent and anomalous outcomes. Minimising less reliable aspects and maximising the effectiveness of all components and interactions is likely to enhance selection practices for surgical training.

Selection constituency

Candidates aim to present themselves to their utmost advantage to maximise the likelihood that they will be selected. They desire to excel in selection criteria, to demonstrate the extent of their knowledge, skills and abilities and accentuate their alignment with those sought by selectors; candidates also desire to perform at their best, to outperform other candidates and to influence selectors' judgements in their favour. Those responsible for surgical selection may consider the possibility that this 'impression-making' is at odds with candidates' normal demeanour; it may therefore seem apt to set criteria or to pose questions to beguile candidates to reveal their 'true' selves and to extrapolate from candidates' responses their 'likely' behaviour in other circumstances. However, unless criteria and questions have been shown to have acceptable validity and reliability, such inferences should be treated with scepticism. Those responsible for selection should be wary of embracing pseudo-scientific solutions, or taking leaps of logic such as attributing candidate responses and behaviour to unconfirmed causes or drawing unjustifiable inferences about candidates' skills, attitudes or aptitude. Selectors seek to discriminate between closely matched candidates, to ascertain individuals' performance against selection criteria, to anticipate their future performance, to maximise the likelihood that those selected will successfully complete training and become proficient surgeons and to minimise the prospects that they would leave the program before completion or become inept surgeons.

Intangible and concrete aspects of selection

Stated and 'unstated' selection criteria affect the fairness of selection; adhering to the stated, overt criteria and minimising intangible, or unstated selection criteria and processes improve procedural justice. Candidates who assume that they will be assessed against stated and unstated selection criteria, are likely to try to second-guess what the unstated criteria might be. This could distract candidates from addressing the stated selection criteria, detract from their performance, increase their anxiety and frustration, and diminish their satisfaction with the selection process. Attempting to demonstrate knowledge, skills and attributes that a candidate considers important, but which are not authorised selection criteria, could take candidates' time and energy that could otherwise be spent addressing genuine, confirmed selection requirements. This has the potential to reduce the efficiency of the process and compromise outcomes. As mentioned above, selectors at times may attempt to second-guess the suitability of candidates for training. Accountability constraints regulating RACS' SET program result in selectors being less likely to overtly circumvent selection processes than to deviate from good practice, (inadvertently or by consciously 'bending the rules', for example, by making unsupported assumptions or asking inappropriate questions). Although there may always be individualistic, or even uncontrolled elements in selection, clear statements of dependable processes and reliable selection criteria, coupled with demonstrated adherence to these, reduce the scope for unsubstantiated conjecture and increase candidate and selector confidence and perceptions of procedural fairness.

Selection instruments and activities

This study has identified that the CV—scoring academic achievements, research undertakings and performance of non-surgical activities—is an ineffectual predictor of performance in the SET program. The RR—scoring similar attributes to the ETA and assessing past performance observed over time—is the strongest predictor of future performance. The multi-station Int—providing multiple instances of performance in multiple domains to multiple assessors, (albeit within a single-day timeframe)—shows some capacity to predict future performance, however, has some of the limitations of a one-off examination, as performance is assessed on a single occasion (day) that might vary from the candidate's usual performance. Total Sel—combining scores in the three selection instruments—also shows capacity to predict later performance.

These results, differing as they do from other researchers' findings—particularly regarding the RR—demonstrate that implementation procedures are critically important components of selection. It is hypothesised that the local implementation of the RR in selection to GS SET—whereby 'referees' were not directly chosen by candidates, but were appointed by the those responsible for GS selection—is an important factor in the findings of this study. This referee appointment process makes such referee reports more likely to represent impartial assessments of candidate performance.

The findings of this study have shown that performance in RACS GS selection partially predicts performance in assessments during training. This indicates that RACS GS selection tools are performing moderately well—evidenced by correlations between RR and later assessments and between Int and later assessments—but that the selection instruments and processes could be modified to maximise their effectiveness. Differences observed between alignments of selection and assessment performance in each year in the study tend to confound the results. The practice of annually modifying selection practices on anecdotal or limited evidence is not recommended.

Although unique, RACS' SET program forms part of a global 'network' of surgical training. Surgical training programs worldwide face many similar issues. Reviewing others' responses to selection, training and assessment, testing and adapting them to local circumstances may increase the range of selection options available to SET, without overstretching local resources. Regular review of RACS' and others' selection and assessment techniques, to ascertain best practice and identify new developments in selection would enable RACS and GS to evaluate the suitability of others' endeavours for SET and the social and regulatory environments in which it operates. Piloting new instruments and activities in the local context to ascertain their efficacy would provide evidence for any ensuing recommendations.

Establishing the attributes that RACS GS considers beneficial to training may assist this specialty to revise and determine selection instruments and processes. This study found no evidence that RACS GS has identified attributes that might be advantageous to GS training or practice. The nine RACS competencies could provide a basis for such activity. Clarifying aspects of these competencies—as expressions of knowledge, skills and attributes—with regard to specialty requirements, identifying the standards of performance required at entry level and formalising statements of these requirements is likely to benefit selectors and candidates alike. Alignment of selection instruments with identified attributes is one outcome to be sought. Situational judgement tests have been identified elsewhere as addressing identified attributes. Entrustable professional activities have been identified as addressing tasks undertaken in the early stages of training. Since this study commenced, GSA has introduced assessments of specified procedural skills and professional capabilities—similar to EPAs—as a selection component (*2016 Selection for 2017 Intake Guidelines*, 2016). A future study could observe relationships between performance in procedural skills and in later assessments.

Phases, events, interstices, interactions and processes

The multiple interactions between selection components impact on selection events, processes, influences and outcomes. The model, Influences on selection (Literature Review chapter Figure 3), identifies the main components engaged in selection to RACS' surgical training program. A considerable challenge in selection activities is to strike an appropriate balance between the interactions and effects of these various components. The regulatory and social frameworks, the goals of selection, the validity, reliability, fairness and feasibility of the instruments and processes and human agency interact to determine who will be selected into the training program. Greater or lesser emphasis on any of these components will affect the processes and the outcomes. Each of these aspects can be influenced by, and influences, the other components: regulatory and social frameworks influence the degree to which procedural justice and fairness are supported in selection processes. Social frameworks and human agency affect the identification of knowledge, skills and attributes considered to be important in training and beyond. Human agency and choice of selection instruments and processes determine how effectively procedural justice goals are implemented in practice and how effectively the instruments address identified knowledge, skills and attributes. The choice of selection instruments is influenced by all these factors and the perceived 'inherent' validity, reliability, fairness and feasibility of the selection instruments and processes under review.

To elaborate on the impact of considerations pertaining to procedural justice, selectors may struggle, for example, with the restrictions of consciously basing selection decisions on limited evidence rather than on using their intuition or 'sub-conscious perceptions' (revealed in statements such as, 'I know it when I see it—this person will be a great trainee and surgeon'); however, obligations concerning accountability, transparency and procedural justice, require evidence-based judgements (Eva and Reiter, 2004). Therefore, to ensure that accountability requirements are met and selectors are satisfied that they have sufficient, relevant evidence on which to base their judgements, great importance rests on the efficacy of selection instruments and processes to provide adequate, appropriate, meaningful, valid and reliable evidence. Additional considerations with evidence-based selection include addressing selectors' concerns about the efficacy of selection instruments and processes, training selectors in their use and informing them about the attributes identified as most suitable for the training program.

Similarly, candidates can become frustrated if they consider that they have not had opportunities to adequately present their experiences, knowledge, skills and attributes to selectors. Candidates with strong opinions about what they would like to present to selectors will be dissatisfied with selection methods that do not allow them to do so. However, candidates' perceptions and experiences of selection are ordinarily confined to their personal perspectives, limiting their ability to fairly judge the whole process or the overarching concepts. For candidate satisfaction to be enhanced, selection tools and processes must be perceived as procedurally fair and that the activities by which their performance is judged are relevant to surgical training. Again, increasing procedural justice and addressing candidates' concerns will involve informing them about what is being sought in selection and the efficacy of instruments and processes.

Those responsible for selection grapple with ascertaining the likelihood of candidates' current and future performance. In selection to SET, evidence for judgements is gleaned from candidates' reports of their experiences (CV) and responses to questions (Int), from assessments of their performance (RRs) and from direct observation of their performance (Int). Limitations of, and tensions between interacting elements—human agency, requirement specifications, instruments, processes, influences—in selection to surgical training, mean that the outcomes of selection can never be certain. This study has identified the principal elements in selection to surgical training, discussed influences on, and interrelationships between, the elements, and has reviewed connections between selection and surgical training assessments. Awareness of limitations, advantages and purposes of selection, training and assessment instruments and processes; recognition of selection, training and assessment objectives; adherence to procedures and regular review of instruments and processes, coupled with caution in changing these measures will contribute to the efficacy of all these elements in the continuum of surgical training. Maximising the effectiveness, efficiency and linkages between SET selection and assessments is likely to engender trainees and surgeons who manifest the RACS' values and vision, and ensure ongoing quality surgical care to the Australian and New Zealand communities.

REFERENCES

2013 Census – Major ethnic groups in New Zealand. (2015). Stats.govt.nz. Retrieved 15 September 2017, from http://www.stats.govt.nz/Census/2013-census/profile-and-summary-reports/infographic-culture-identity.aspx

2016 Selection for 2017 Intake Guidelines. (2016) (1st ed.). Melbourne. Retrieved from http://www.generalsurgeons.com.au/media/files/Education%20and%20Training/Selection/GDE%202015-12-09%20Final.pdf

2071.0 - Census of Population and Housing: Reflecting Australia - Stories from the Census, 2016. (2017). *Abs.gov.au*. Retrieved 5 October 2017, from http://www.abs.gov.au/ausstats/abs@.nsf/mf/2071.0

3412.0 - Migration, Australia, 2014-15. (2016). *Abs.gov.au*. Retrieved 15 September 2017, from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3412.0Main+Features12014-15

About RACS | Royal Australasian College of Surgeons. (2017). Surgeons.org. Retrieved 2 September 2017, from https://surgeons.org/about/

Accreditation and recognition. (2017). *Australian Medical Council*. Retrieved 13 March 2017, from https://www.amc.org.au/accreditation

Adam, J., Bore, M., McKendree, J., Munro, D., & Powis, D. (2012). Can personal qualities of medical students predict in-course examination success and professional behaviour? An exploratory prospective cohort study. *BMC Medical Education*, *12*(1), 69. http://dx.doi.org/10.1186/1472-6920-12-69

Adam, J., Dowell, J., & Greatrix, R. (2011). Use of UKCAT scores in student selection by UK medical schools, 2006-2010. *BMC Medical Education*, *11*(1), 98. http://dx.doi.org/https://doi.org/10.1186/1472-6920-11-98

Albanese, M., Snow, M., Skochelak, S., Huggett, K., & Farrell, P. (2003). Assessing Personal Qualities in Medical School Admissions. *Academic Medicine*, *78*(3), 313-321. http://dx.doi.org/10.1097/00001888-200303000-00016

Alterman, D., Jones, T., Heidel, R., Daley, B., & Goldman, M. (2011). The Predictive Value of General Surgery Application Data for Future Resident Performance. Journal of Surgical Education, 68(6), 513-518. http://dx.doi.org/10.1016/j.jsurg.2011.07.007

amc.org.au,. (2015). *Australian Medical Council » Accreditation Reports*. Retrieved 26 September 2015, from http://www.amc.org.au/accreditation/medical-education/accreditation-reports

Anastakis, D., Hamstra, S., & Matsumoto, E. (2000). Visual-spatial abilities in surgical training. *The American Journal of Surgery*, *179*(6), 469-471. http://dx.doi.org/10.1016/s0002-9610(00)00397-4

Anderson, L. W., and Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives: Complete edition, New York: Longman. pp. 67-68

Andriole, D., Jeffe, D., & Whelan, A. (2004). What predicts surgical internship performance? *The American Journal of Surgery*, *188*(2), 161--164.

Anton, N., Montero, P., Howley, L., Brown, C., & Stefanidis, D. (2015). What stress coping strategies are surgeons relying upon during surgery?. The American Journal of Surgery, 210(5), 846-851. http://dx.doi.org/10.1016/j.amjsurg.2015.04.002

Argy, F. (2006). *Equality of Opportunity in Australia - Myth and Reality*. Canberra: The Australia Institute.

Armstrong, M. (2006). *A handbook of human resource management practice* (10th ed.). London: Kogan Page.

Arnold, L. (2002). Assessing Professional Behavior: Yesterday, Today, and Tomorrow. Academic Medicine, 77(6), 502-515. http://dx.doi.org/10.1097/00001888-200206000-00006

Arora, S., Sevdalis, N., Nestel, D., Woloshynowych, M., Darzi, A., & Kneebone, R. (2010). The impact of stress on surgical performance: A systematic review of the literature. Surgery, 147(3), 318-330.e6. http://dx.doi.org/10.1016/j.surg.2009.10.007

Arthur, W., Glaze, R., Jarrett, S., White, C., Schurig, I., & Taylor, J. (2014). Comparative evaluation of three situational judgment test response formats in terms of construct-related validity, subgroup differences, and susceptibility to response distortion. *Journal of Applied Psychology*, *99*(3), 535-545. http://dx.doi.org/10.1037/a0035788

Association of American Medical Colleges. (2002). *A guide to the preparation of the medical student performance evaluation*.

Auewarakul, C., Downing, S., Jaturatamrong, U., & Praditsuwan, R. (2005). Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment methods. *Medical Education*, *39*(3), 276-283. http://dx.doi.org/10.1111/j.1365-2929.2005.02090.x

Australian Health Practitioner Regulation Agency - Legal Practice Notes - LPN 17. (2013). *ahpra.gov.au*. Retrieved 1 December 2015, from https://www.ahpra.gov.au/Publications/legal-practice-notes.aspx

Australian Health Practitioner Regulation Agency - Registration Process. (2017). *Ahpra.gov.au*. Retrieved 5 October 2017, from http://www.ahpra.gov.au/Registration/Registration-Process.aspx

Australian Medical Council » Accredited medical schools. (2017). *Amc.org.au*. Retrieved 4 October 2017, from https://www.amc.org.au/accreditation/primary-medical-education/schools

Australian Medical Council » Assessing primary medical education. (2017). *Amc.org.au*. Retrieved 4 October 2017, from https://www.amc.org.au/accreditation/primary-medical-education

Australian Medical Council (AMC),. (2007). *Standards for Assessment and Accreditation of Specialist Medical Education Programs and Professional Development programs by the Australian Medical Council 2007*. Kingston, ACT, Australia: Australian Medical Council.

Baldwin, D. (2003). Toward a theory of professional development: framing humanism at the core of good doctoring and good pedagogy. In Arnold P. Gold Foundation overcoming the barriers to Sustaining Humanism in Medicine Symposium (pp. 6-10). New York: Arnold P. Gold Foundation.

Bann, S., & Darzi, A. (2005). Selection of individuals for training in surgery. *The American Journal of Surgery*, *190*(1), 98-102. http://dx.doi.org/10.1016/j.amjsurg.2005.04.002

Bann, S., Davis, I., Moorthy, K., Munz, Y., Hernandez, J., & Khan, M. et al. (2005). The reliability of multiple objective measures of surgery and the role of human performance. *The American Journal of Surgery*, *189*(6), 747-752. http://dx.doi.org/10.1016/j.amjsurg.2005.03.020

Bannon, M. (2005). What's happening in postgraduate medical education?. *Archives of Disease in Childhood*, *91*(1), 68-70. http://dx.doi.org/10.1136/adc.2004.070037

Batten, J. (2017). *AMC Accreditation*. Presentation, http://www.surgeons.org/media/25063405/AMC-Accreditation-J-Batten.pdf.

Beard, J. (2007). Assessment of surgical competence. *British Journal of Surgery*, 94(11), 1315-1316. http://dx.doi.org/10.1002/bjs.6048

Beard, J., Marriott, J., Purdie, H., & Crossley, J. (2011). Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. *Health Technology Assessment*, 15(01). http://dx.doi.org/10.3310/hta15010

Beasley, A. (2002). *The mantle of surgery*. Melbourne: Royal Australasian College of Surgeons.

Beasley, S. (2015). Understanding the responsibilities and obligations of the modern paediatric surgeon. *Journal of Pediatric Surgery*, *50*(2), 223-231. http://dx.doi.org/10.1016/j.jpedsurg.2014.11.003

Becoming a Doctor. (2017). *Australian Medical Association*. Retrieved 4 October 2017, from https://ama.com.au/careers/becoming-a-doctor

Bell, R., Banker, M., Rhodes, R., Biester, T., & Lewis, F. (2007). Graduate Medical Education in Surgery in the United States. *Surgical Clinics of North America*, *87*(4), 811-823. doi:10.1016/j.suc.2007.06.005

Bell, R., Fann, S., Morrison, J., & Lisk, J. (2011). Determining personal talents and behavioral styles of applicants to surgical training: a new look at an old problem, part I. *Journal of Surgical Education*, *68*(6), 534--541.

Bell, R., Fann, S., Morrison, J., & Lisk, J. (2012). Determining personal talents and behavioral styles of applicants to surgical training: a new look at an old problem, part II. *Journal of Surgical Education*, *69*(1), 23–29.

Bergman, M., Drasgow, F., Donovan, M., Henning, J., & Juraska, S. (2006). Scoring Situational Judgment Tests: Once You Get the Data, Your Troubles Begin. *International Journal of Selection and Assessment*, *14*(3), 223-235. http://dx.doi.org/10.1111/j.1468-2389.2006.00345.x

Birks, D., & Palermo, C. (2008). Surgical Education and Training: Maintaining the competence of supervisors. *ANZ Journal of Surgery*, *78*(9), 737-738. http://dx.doi.org/10.1111/j.1445-2197.2008.04639.x

Bishawi, M. & Pryor, A. (2014). Should technical aptitude evaluation become part of resident selection for surgical residency?. *Surgical Endoscopy*, 28(10), 2761-2762. http://dx.doi.org/10.1007/s00464-014-3715-3

Blader, S., & Tyler, T. (2003). What constitutes fairness in work settings? A four-component model of procedural justice. *Human Resource Management Review*, *13*(1), 107--126.

Blake, P., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: the knowledge–behavior gap. *Trends in Cognitive Sciences*, 18(11), 559-561. http://dx.doi.org/10.1016/j.tics.2014.08.003

Bloom, B. (1956). *Taxonomy of educational objectives*. N.Y.: Longmans, Green.

Bls.gov,. (2015). *Physicians and Surgeons: Occupational Outlook Handbook: U.S. Bureau of Labor Statistics*. Retrieved 3 October 2015, from http://www.bls.gov/ooh/healthcare/physicians-and-surgeons.htm#tab-4

Bore, M., Munro, D., & Powis, D. (2009). A comprehensive model for the selection of medical students. *Medical Teacher*, *31*(12), 1066--1072.

Breaugh, J. (2008). Employee recruitment: Current knowledge and important areas for future research. *Human Resource Management Review*, *18*(3), 103-118. http://dx.doi.org/10.1016/j.hrmr.2008.07.003

Buckley, C., Kavanagh, D., Nugent, E., Ryan, D., Traynor, O., & Neary, P. (2014). The impact of aptitude on the learning curve for laparoscopic suturing. *The American Journal of Surgery*, 207(2), 263-270. http://dx.doi.org/10.1016/j.amjsurg.2013.08.037

Buhr, H., Gröne, J., & Ritz, J. (2012). Commentary on “Trainee satisfaction in surgical residency programs: Modern management tools ensure trainee motivation and success” by von Websky et al. *Surgery*, *152*(5), 807-808. doi:10.1016/j.surg.2012.08.049

Burgess, A., Roberts, C., Clark, T., & Mossman, K. (2014). The social validity of a national assessment centre for selection into general practice training. *BMC Medical Education, 14*(1). doi:10.1186/s12909-014-0261-6

Cable, D., & Yu, K. (2006). Managing job seekers' organizational image beliefs: The role of media richness and media credibility. *Journal of Applied Psychology*, *91*(4), 828-840. doi:10.1037/0021-9010.91.4.828

Cameron, J. (1997). William Stewart Halsted – Our surgical heritage. *Annals of Surgery*, *225*(5), 445-458. doi:10.1097/00000658-199705000-00002

Campagna-Vaillancourt, M., Manoukian, J., Razack, S., & Nguyen, L. (2014). Acceptability and reliability of multiple mini interviews for admission to otolaryngology residency. *The Laryngoscope*, *124*(1), 91-96. http://dx.doi.org/10.1002/lary.24122

CanMEDS // About. (2017). Canmeds.royalcollege.ca. Retrieved 6 October 2017, from http://canmeds.royalcollege.ca/en/about

Cappelen, A., Nielsen, U., Tungodden, B., Tyran, J., & Wengström, E. (2015). Fairness is intuitive. *Experimental Economics*, *19*(4), 727-740. http://dx.doi.org/10.1007/s10683-015-9463-y

Carmichael, K., Westmoreland, J., Thomas, J., and Patterson, R. (2005). Relation of Residency Selection Factors to Subsequent Orthopaedic In-training Examination Performance. *Southern Medical Journal*, *98*(5), 528-532. doi:10.1097/01.smj.0000157560.75496.cb

Carms.ca,. (2015). *Canadian Resident Matching Service - Service canadien de Jumelage des résidents | CaRMS*. Retrieved 3 October 2015, from http://www.carms.ca

Carroll, S., Kennedy, A., Traynor, O., & Gallagher, A. (2009). Objective assessment of surgical performance and its impact on a national selection programme of candidates for higher surgical training in plastic surgery. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, *62*(12), 1543-1549. http://dx.doi.org/10.1016/j.bjps.2008.06.054

CIPD - Diversity in the workplace: an overview - Factsheets. (2015). Cipd.co.uk. Retrieved 16 April 2016, from http://www.cipd.co.uk/hr-resources/factsheets/diversity-workplace-overview.aspx

Clark, W., & Maas, R. (2013). *Socio-spatial mobility and neighbourhood outcomes in Australia: How "fluid" is Australian society?*. *Melbourne Institute*. Retrieved 5 January 2016, from https://www.melbourneinstitute.com/downloads/conferences/HILDA_2013/HILDA_2013_papers/Clark,%20William_final%20paper.pdf

Cleland, J., Dowell, J., McLachlan, J., Nicholson, S., & Patterson, F. (2012). *Identifying best practice in the selection of medical students (literature review and interview survey)*. General Medical Council. Retrieved from http://www.gmc-uk.org/Identifying_best_practice_in_the_selection_of_medical_students.pdf_51119804.pdf

Cleland, J., Patterson, F., Dowell, J., and Nicholson, S. (2014). How can greater consistency in selection between medical schools be encouraged? A mixed-methods programme of research that examines and develops the evidence base. London: University of Aberdeen; University of Dundee; Queen Mary University of London; Work Psychology Group.

Coebergh, J. (2003). Dutch medical schools abandon selection for lottery system for places. *Student BMJ*, *11*, 138.

Cohen-Schotanus, J., Muijtjens, A., Reinders, J., Agsteribbe, J., van Rossum, H., & van der Vleuten, C. (2006). The predictive validity of grade point average scores in a partial lottery medical school admission system. *Medical Education, 40*(10), 1012-1019. doi:10.1111/j.1365-2929.2006.02561.x

Collins, J. The selection and work-based assessment of surgical trainees - A discussion document. January 2007. Melbourne: Royal Australasian College of Surgeons.

Collins, J., Civil, I., Sugrue, M., Balogh, Z., & Chehade, M. (2008). Surgical Education and Training in Australia and New Zealand. *World Journal of Surgery*, *32*(10), 2138-2144. http://dx.doi.org/10.1007/s00268-008-9680-z

Collins, J., Gough, I., Civil, I., & Stitz, R. (2007). A New Surgical Education and Training Programme. *ANZ Journal of Surgery*, *77*(7), 497-501. http://dx.doi.org/10.1111/j.1445-2197.2007.04171.x

Commonwealth Department of Health and Family Services. (1998). *Trainee selection in Australian medical colleges*. Canberra: Commonwealth Department of Health and Family Services.

Competencies | Royal Australasian College of Surgeons. (2017). *Surgeons.org*. Retrieved 19 February 2017, from http://www.surgeons.org/becoming-a-surgeon/surgical-education-training/competencies/

Cone Jr, J., Byrum, C., Payne, W., & Smith Jr, D. (2012). A Novel Adjuvant to the Resident Selection Process: the Hartman Value Profile. *Eplasty*, *12*.

Crossley, J., & Jolly, B. (2011). Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Medical Education*, *46*(1), 28-37. doi:10.1111/j.1365-2923.2011.04166.x

Crossley, J., Davies, H., Humphris, G., & Jolly, B. (2002). Generalisability: a key to unlock professional assessment. *Medical Education*, *36*(10), 972-978. http://dx.doi.org/10.1046/j.1365-2923.2002.01320.x

Crossley, J., Humphris, G., & Jolly, B. (2002). Assessing health professionals. *Medical Education*, 36(9), 800-804. http://dx.doi.org/10.1046/j.1365-2923.2002.01294.x

Cruess, R., Cruess, S., & Steinert, Y. (2009). Teaching medical professionalism. Cambridge: Cambridge University Press.

Culbertson, S., Weyhrauch, W., & Huffcutt, A. (2017). A tale of two formats: Direct comparison of matching situational and behavior description interview questions. *Human Resource Management Review*, *27*(1), 167-177. http://dx.doi.org/10.1016/j.hrmr.2016.09.009

Cullen, M., Reed, D., Halvorsen, A., Wittich, C., Kreuziger, L., & Keddis, M. et al. (2011). Selection criteria for internal medicine residency applicants and professionalism ratings during internship, *86*(3), 197--202.

Cullen, M., Sackett, P., & Lievens, F. (2006). Threats to the Operational Use of Situational Judgment Tests in the College Admission Process. *International Journal of Selection and Assessment*, *14*(2), 142-155. http://dx.doi.org/10.1111/j.1468-2389.2006.00340.x

Cuschieri, A., Francis, N., Crosby, J., & Hanna, G. (2001). What do master surgeons think of surgical competence and revalidation? *The American Journal of Surgery*, 182(2), 110-116. http://dx.doi.org/10.1016/s0002-9610(01)00667-5

Davis, M. & Ponnamperuma, G. (2005). Portfolio assessment. *Journal of Veterinary Medical Education*, 32(3), 279-284. http://dx.doi.org/10.3138/jvme.32.3.279

Deary, I. (1998). Differences in mental abilities. BMJ, 317(7174), 1701-1703. http://dx.doi.org/10.1136/bmj.317.7174.1701

Delbridge, A., Bernard, J., Blair, D., Butler, S., Peters, P., & Yallop, C. (2003). *The Macquarie Dictionary* (3rd ed., pp. 380, 1320, 1871). St Leonards, N.S.W., Australia: Macquarie Library Pty Ltd.

Department of Health and Human Services, (2016). Delivering for diversity - Cultural diversity plan 2016-2019 (1st ed.). Melbourne: Victorian Government Department of Health and Human Services. Retrieved from https://www2.health.vic.gov.au/about/publications/policiesandguidelines/dhhs-delivering-for-diversity-cultural-diversity-plan-2016-19

Dijkstra, J., Galbraith, R., Hodges, B., McAvoy, P., McCrorie, P., & Southgate, L. et al. (2012). Expert validation of fit-for-purpose guidelines for designing programmes of assessment. BMC Medical Education, 12(1), 20. http://dx.doi.org/10.1186/1472-6920-12-20

Dirschl, D. (2002). Scoring of orthopaedic residency applicants: is a scoring system reliable?. *Clinical Orthopaedics and Related Research*, 399, 260--264.

Dirschl, D., Dahners, L., Adams, G., Crouch, J., & Wilson, F. (2002). Correlating selection criteria with subsequent performance as residents. *Clinical Orthopaedics and Related Research*, 399, 265--274.

Dodson, M., Crotty, B., Prideaux, D., Carne, R., Ward, A., & De Leeuw, E. (2009). The multiple mini-interview: how long is long enough?. *Medical Education*, *43*(2), 168--174.

Dore, K., Kreuger, S., Ladhani, M., Rolfson, D., Kurtz, D., & Kulasegaram, K. et al. (2010). The reliability and acceptability of the multiple mini-interview as a selection instrument for postgraduate admissions. *Academic Medicine*, *85*(10), 60--63.

Dowell, J., Lumsden, M., Powis, D., Munro, D., Bore, M., Makubate, B., & Kumwenda, B. (2011). Predictive validity of the personal qualities assessment for selection of medical students in Scotland. *Medical Teacher*, 33(9), e485-e488. http://dx.doi.org/10.3109/0142159x.2011.599448

Dowell, J., Lynch, B., Till, H., Kumwenda, B., & Husbands, A. (2012). The multiple mini-interview in the UK context: 3 years of experience at Dundee. *Medical Teacher*, *34*(4), 297--304.

Downing, S. (2003). Validity: on the meaningful interpretation of assessment data. *Medical Education*, 37(9), 830-837. http://dx.doi.org/10.1046/j.1365-2923.2003.01594.x

Downing, S. (2004). Reliability: on the reproducibility of assessment data. *Medical Education*, 38(9), 1006-1012. http://dx.doi.org/10.1111/j.1365-2929.2004.01932.x

Edwards, D., Friedman, T., & Pearce, J. (2013). Same admissions tools, different outcomes: a critical perspective on predictive validity in three undergraduate medical schools. *BMC Medical Education*, 13(1), 173.

Edwards, J. (2008). Person–Environment Fit in Organizations: An Assessment of Theoretical Progress. *The Academy of Management Annals*, *2*(1), 167-230. doi:10.1080/19416520802211503

Elfenbein, D., Sippel, R., McDonald, R., Watson, T., Scarborough, J., & Migaly, J. (2015). Faculty evaluations of resident medical knowledge: can they be used to predict American Board of Surgery In-Training Examination performance?. *The American Journal of Surgery*, 209(6), 1095-1101. http://dx.doi.org/10.1016/j.amjsurg.2014.08.042

Ellis, A., West, B., Ryan, A., & DeShon, R. (2002). The use of impression management tactics in structured interviews: A function of question type?. *Journal of Applied Psychology*, *87*(6), 1200-1208. http://dx.doi.org/10.1037//0021-9010.87.6.1200

Emery, J., & Bell, J. (2009). The predictive validity of the BioMedical Admissions Test for pre-clinical examination performance. *Medical Education*, *43*(6), 557--564.

Emery, J., Bell, J., & Vidal Rodeiro, C. (2011). The BioMedical Admissions Test for medical student selection: Issues of fairness and bias. *Medical Teacher*, *33*(1), 62-71.

Ericsson, K. (2015). Acquisition and Maintenance of Medical Expertise. *Academic Medicine*, 90(11), 1471-1486. http://dx.doi.org/10.1097/acm.0000000000000939

Eva, K. (2014). Moving beyond childish notions of fair and equitable. *Medical Education, 49*(1), 1-3. doi:10.1111/medu.12640

Eva, K. & Macala, C. (2014). Multiple mini-interview test characteristics: 'tis better to ask candidates to recall than to imagine. *Medical Education*, *48*(6), 604-613. http://dx.doi.org/10.1111/medu.12402

Eva, K., & Reiter, H. (2004). Where Judgement Fails: Pitfalls in the Selection Process for Medical Personnel. *Advances in Health Sciences Education*, *9*(2), 161-174. http://dx.doi.org/10.1023/b:ahse.0000027479.14637.6c

Eva, K., Reiter, H., Rosenfeld, J., and Norman, G. (2004). The Ability of the Multiple Mini-Interview to Predict Preclerkship Performance in Medical School. *Academic Medicine*, *79*(Supplement), S40-S42. doi:10.1097/00001888-200410001-00012

Eva, K., Reiter, H., Rosenfeld, J., Trinh, K., Wood, T., & Norman, G. (2012). Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. *JAMA*, *308*(21), 2233--2240.

Eva, K., Reiter, H., Trinh, K., Wasi, P., Rosenfeld, J., & Norman, G. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education*, 43(8), 767--775.

Eva, K., Rosenfeld, J., Reiter, H., & Norman, G. (2004). An admissions OSCE: the multiple mini-interview. *Medical Education*, *38*(3), 314-326. http://dx.doi.org/10.1046/j.1365-2923.2004.01776.x

Evgeniou, E., Peter, L., Tsironi, M., & Iyer, S. (2013). Assessment methods in surgical training in the United Kingdom. *Journal of Educational Evaluation For Health Professions*, *10*, 2. doi:10.3352/jeehp.2013.10.2

Farah, S., Winter, M., & Smith, J. (2011). Current medical students' understanding of Surgical Education and Training. *ANZ Journal of Surgery*, 82(1-2), 12-16. http://dx.doi.org/10.1111/j.1445-2197.2011.05944.x

Farkas, D., Nagpal, K., Curras, E., Shah, A., & Cosgrove, J. (2012). The Use of a Surgery-Specific Written Examination in the Selection Process of Surgical Residents. *Journal of Surgical Education*, 69(6), 807-812. http://dx.doi.org/10.1016/j.jsurg.2012.05.011

Forehand, M (The University of Georgia) in Orey 2010 Emerging perspectives on learning teaching and technology pp 41-43

Forsythe, A. & Johnson, S. (2016). Thanks, but no-thanks for the feedback. Assessment and Evaluation in Higher Education, 1-10. http://dx.doi.org/10.1080/02602938.2016.1202190

Foster, K., Neidert, G., Brubaker-Rimmer, R., Artalejo, D., & Caruso, D. (2010). A Psychological Profile of Surgeons and Surgical Residents. *Journal of Surgical Education*, 67(6), 359-370. http://dx.doi.org/10.1016/j.jsurg.2010.07.007

Fromm, E. (2002). *The fear of freedom*. London: Routledge.

Fromme, H., Karani, R., & Downing, S. (2009). Direct Observation in Medical Education: A Review of the Literature and Evidence for Validity. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, 76(4), 365-371. http://dx.doi.org/10.1002/msj.20123

Gallagher, A., Leonard, G., & Traynor, O. (2009). Role and feasibility of psychomotor and dexterity testing in selection for surgical training. *ANZ Journal of Surgery*, 79(3), 108--113.

Gallagher, A., Neary, P., Gillen, P., Lane, B., Whelan, A., Tanner, W., & Traynor, O. (2008). Novel method for assessment and selection of trainees for higher surgical training in general surgery. *ANZ Journal of Surgery*, 78(4), 282--290.

Gallagher, A., O'Sullivan, G., Neary, P., Carroll, S., Leonard, G., Bunting, B., & Traynor, O. (2014). An objective evaluation of a multi-component, competitive, selection process for admitting surgeons into higher surgical training in a national setting. *World Journal of Surgery*, 38(2), 296--304.

Gallagher, A., Ritter, E., & Satava, R. (2003). Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surgical Endoscopy*, 17(10), 1525-1529. http://dx.doi.org/10.1007/s00464-003-0035-4

Gardner, A., Ritter, E., Paige, J., Ahmed, R., Fernandez, G., & Dunkin, B. (2016). Simulation-Based Selection of Surgical Trainees: Considerations, Challenges, and Opportunities. *Journal of The American College Of Surgeons*, *223*(3), 530-536. http://dx.doi.org/10.1016/j.jamcollsurg.2016.05.021

Gardner, A., Steffes, C., Nepomnayshy, D., Nicholas, C., Widmann, W., & Fitzgibbons, S. et al. (2017). Selection bias: Examining the feasibility, utility, and participant receptivity to incorporating simulation into the general surgery residency selection process. *The American Journal of Surgery*, *213*(6), 1171-1177. http://dx.doi.org/10.1016/j.amjsurg.2016.09.029

Geffen, L. (2014). A brief history of medical education and training in Australia. *The Medical Journal of Australia*, *201*(1), 19-22. doi:10.5694/mja14.00118

Geissler, J., VanHeest, A., Tatman, P., & Gioe, T. (2013). Aggregate Interview Method of Ranking Orthopedic Applicants Predicts Future Performance. *Orthopedics*, 36(7), e966-e970. http://dx.doi.org/10.3928/01477447-20130624-30

Generalsurgeons.com.au,. (2016). *General Surgeons Australia - Selection*. Retrieved 6 January 2016, from http://www.generalsurgeons.com.au/education-and-training/selection

Gilbart, M., Cusimano, M., & Regehr, G. (2001). Evaluating surgical resident selection procedures. *The American Journal of Surgery*, *181*(3), 221-225. http://dx.doi.org/10.1016/s0002-9610(01)00550-5

Gilliland, S. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, *18*(4), 694--734.

Gipps, C. and Stobart, G. (2009). Fairness in Assessment. In C. Wyatt-Smith & J. Cumming, Educational assessment in the 21st century: connecting theory and practice (1st ed., pp. 105-118). Dordrecht; New York: Springer. Retrieved from http://Full-text via Springer English/International eBooks 2009 - Full Set

Girotti, J., Park, Y., & Tekian, A. (2014). Ensuring a fair and equitable selection of students to serve society's health care needs. *Medical Education*, *49*(1), 84-92. http://dx.doi.org/10.1111/medu.12506

Glick, S. (2000). Selection for entry to medicine and specialist training. *Medical Teacher*, 22(5), 443-447. http://dx.doi.org/10.1080/01421590050110678

Go, P., Klaassen, Z., & Chamberlain, R. (2012). Residency selection: do the perceptions of US programme directors and applicants match?. *Medical Education*, *46*(5), 491-500. doi:10.1111/j.1365-2923.2012.04257.x

Gough, M., & Bell, J. (1989). Introducing aptitude testing into medicine. *BMJ*, *298*(6679), 975-976. doi:10.1136/bmj.298.6679.975

Graduate Entry Medical Schools Admission System,. (2015). *Applications | GEMSAS. Gemsas.edu.au*. Retrieved 4 October 2015, from http://www.gemsas.edu.au/applications/

Grantcharov, T., & Reznick, R. (2009). Training tomorrow's surgeons: what are we looking for and how can we achieve it?. *ANZ Journal of Surgery*, 79(3), 104--107.

Griffin, B. & Wilson, I. (2010). Interviewer bias in medical student selection. *The Medical Journal of Australia*, *193*(6), 343--346.

Griffin, B., Carless, S., & Wilson, I. (2012). The effect of commercial coaching on selection test performance. *Medical Teacher*, *35*(4), 295--300.

Griffin, P., Wu, M., & Zoanetti, N. (2004). *Rasch Model Scaling and Reporting for the Royal Australasian College of Surgeons Basic Surgical Sciences MCQ Examination* (pp. 1-21). Melbourne: Assessment Research Centre, The University of Melbourne. Retrieved from https://findanexpert.unimelb.edu.au/display/publication32845

Hall, J., Ellis, C., & Hamdorf, J. (2003). Surgeons and cognitive processes. British Journal of Surgery, 90(1), 10-16. http://dx.doi.org/10.1002/bjs.4020

Harris, B., Walsh, J., & Lammy, S. (2015). UK medical selection: lottery or meritocracy? *Clinical Medicine*, *15*(1), 40-46. http://dx.doi.org/10.7861/clinmedicine.15-1-40

Hay, M., Mercer, A., Lichtwark, I., Tran, S., Hodgson, W., & Aretz, H. et al. (2017). Selecting for a sustainable workforce to meet the future healthcare needs of rural communities in Australia. *Advances in Health Sciences Education*, *22*(2), 533-551. http://dx.doi.org/10.1007/s10459-016-9727-0

Hern Jr, H., Alter, H., Wills, C., Snoey, E., & Simon, B. (2013). How prevalent are potentially illegal questions during residency interviews?. *Academic Medicine*, *88*(8), 1116--1121.

Hofmeister, M., Lockyer, J., & Crutcher, R. (2008). The acceptability of the Multiple Mini Interview for resident selection. *Family Medicine*, *40*(10), 734-740.

Holmboe, E., Huot, S., Chung, J., Norcini, J., & Hawkins, R. (2003). Construct Validity of the MiniClinical Evaluation Exercise (MiniCEX). *Academic Medicine*, 78(8), 826-830. http://dx.doi.org/10.1097/00001888-200308000-00018

Husbands, A., Rodgerson, M., Dowell, J., & Patterson, F. (2015). Evaluating the validity of an integrity-based situational judgement test for medical school admissions. *BMC Medical Education*, *15*(1). http://dx.doi.org/10.1186/s12909-015-0424-0

Iobst, W., Sherbino, J., ten Cate, O., Richardson, D., Dath, D., & Swing, S. et al. (2010). Competency-based medical education in postgraduate medical education. *Medical Teacher, 32*(8), 651-656. doi:10.3109/0142159x.2010.500709

Jefferis, T. (2007). Selection for specialist training: what can we learn from other countries?. *BMJ*, 334(7607), 1302-1304. http://dx.doi.org/10.1136/bmj.39238.447338.ad

Karantonis, A., & Sireci, S. (2006). The Bookmark Standard-Setting Method: A Literature Review. *Educational Measurement: Issues and Practice*, 25(1), 4-12. http://dx.doi.org/10.1111/j.1745-3992.2006.00047.x

Katsufrakis, P., Uhler, T., & Jones, L. (2016). The Residency Application Process. *Academic Medicine*, 1. http://dx.doi.org/10.1097/acm.0000000000001411

Kelly, M., Dowell, J., Husbands, A., Newell, J., O'Flynn, S., & Kropmans, T. et al. (2014). The fairness, predictive validity and acceptability of multiple mini interview in an internationally diverse student population- a mixed methods study. *BMC Medical Education*, *14*(1), 267-280. http://dx.doi.org/10.1186/s12909-014-0267-0

Kelz, R., Mullen, J., Kaiser, L., Pray, L., Shea, G., & Drebin, J. et al. (2010). Prevention of Surgical Resident Attrition by a Novel Selection Strategy. *Annals of Surgery*, *25*(3), 537-543. doi:10.1097/sla.0b013e3181f27a50

Kenny, S., McInnes, M., & Singh, V. (2013). Associations between residency selection strategies and doctor performance: a meta-analysis. *Medical Education*, *47*(8), 790-800. http://dx.doi.org/10.1111/medu.12234

Khan, K., & Ramachandran, S. (2012). Conceptual framework for performance assessment: Competency, competence and performance in the context of assessments in healthcare – Deciphering the terminology. *Medical Teacher, 34*(11), 920-928. doi:10.3109/0142159x.2012.722707

Klotz, A., Motta Veiga, S., Buckley, M., & Gavin, M. (2013). The role of trustworthiness in recruitment and selection: A review and guide for future research. *Journal of Organizational Behavior*, *34*(S1), 104--119.

Knorr, M. & Hissbach, J. (2014). Multiple mini-interviews: same concept, different approaches. *Medical Education*, *48*(12), 1157-1175. http://dx.doi.org/10.1111/medu.12535

Koczwara, A., Patterson, F., Zibarras, L., Kerrin, M., Irish, B., & Wilkinson, M. (2012). Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Medical Education*, 46(4), 399-408. http://dx.doi.org/10.1111/j.1365-2923.2011.04195.x

Kolk, N., Born, M., & der Flier, H. (2003). The Transparent Assessment Centre: The Effects of Revealing Dimensions to Candidates. *Applied Psychology*, *52*(4), 648-668. doi:10.1111/1464-0597.00156

Kreiter, C., Yin, P., Solow, C., & Brennan, R. (2004). Investigating the Reliability of the Medical School Admissions Interview. *Advances in Health Sciences Education*, *9*(2), 147-159. http://dx.doi.org/10.1023/b:ahse.0000027464.22411.0f

Kron, I., Kaiser, D., Nolan, S., Rudolf, L., Muller, W., & Jones, R. (1985). Can Success in the Surgical Residency be Predicted from Preresidency Evaluation?. *Annals of Surgery*, *202*(6), 694-695. http://dx.doi.org/10.1097/00000658-198512000-00005

Kulatunga-Moruzi, C. & Norman, G. (2002). Validity of Admissions Measures in Predicting Performance Outcomes: The Contribution of Cognitive and Non-Cognitive Dimensions. *Teaching and Learning in Medicine*, 14(1), 34-42. http://dx.doi.org/10.1207/s15328015tlm14019

Kumar, K., Roberts, C., Rothnie, I., du Fresne, C., & Walton, M. (2009). Experiences of the multiple mini-interview: a qualitative analysis. *Medical Education*, *43*(4), 360-367. http://dx.doi.org/10.1111/j.1365-2923.2009.03291.x

Kwakwa, F., & Jonasson, O. (1999). The longitudinal study of surgical residents, 1994 to 1996. *Journal of The American College of Surgeons*, 188(6), 575-585. http://dx.doi.org/10.1016/s1072-7515(99)00044-7

LaGrasso, J., Kennedy, D., Hoehn, J., Ashruf, S., & Przybyla, A. (2008). Selection Criteria for the Integrated Model of Plastic Surgery Residency. *Plastic and Reconstructive Surgery*, *121*(3), 121e-125e. http://dx.doi.org/10.1097/01.prs.0000299456.96822.1b

Leduc, J., Rioux, R., Gagnon, R., Bourdy, C., & Dennis, A. (2016). Impact of sociodemographic characteristics of applicants in multiple mini-interviews. *Medical Teacher*, 1-13. http://dx.doi.org/10.1080/0142159x.2017.1270431

Lemay, J., Lockyer, J., Collin, V., & Brownell, A. (2007). Assessment of non-cognitive traits through the admissions multiple mini-interview. *Medical Education*, *41*(6), 573-579. http://dx.doi.org/10.1111/j.1365-2923.2007.02767.x

Lichtwark, I., Henry, S., Garvey, L., Najm, D., & Hay, M. (2017). Differences in scoring patterns of Situational Judgement Test Scores (SJT) between students of several health professions. In *amee 2017* (#9O5 (652)). Helsinki: The Association for Medical Education in Europe (AMEE). Retrieved from https://amee.org/getattachment/Conferences/AMEE-Past-Conferences/AMEE-2017/AMEE-2017-Abstract-Book.pdf

Lievens, F. (2014). Diversity in medical school admission: insights from personnel recruitment and selection. *Medical Education*, 49(1), 11-14. http://dx.doi.org/10.1111/medu.12615

Lievens, F. & Coetsier, P. (2002). Situational Tests in Student Selection: An Examination of Predictive Validity, Adverse Impact, and Construct Validity. *International Journal of Selection and Assessment*, 10(4), 245-257. http://dx.doi.org/10.1111/1468-2389.00215

Lievens, F., Buyse, T., Sackett, P., & Connelly, B. (2012). The Effects of Coaching on Situational Judgment Tests in High-stakes Selection. *International Journal of Selection and Assessment*, *20*(3), 272-282. http://dx.doi.org/10.1111/j.1468-2389.2012.00599.x

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: a review of recent research. *Personnel Review*, 37(4), 426-441. http://dx.doi.org/10.1108/00483480810877598

Lind, E., & Tyler, T. (1988). *The social psychology of procedural justice*. New York: Plenum Press.

Lind, E., & Van den Bos, K. (2002). When fairness works: Toward a general theory of uncertainty management. *Research in Organizational Behavior*, *24*, 181--223.

Liu, T. & Lu, Y. (2016). Gene–culture interaction and the evolution of the human sense of fairness. *Scientific Reports*, 6, 32483. http://dx.doi.org/10.1038/srep32483

Love, J., Ronan-Bentle, S., Lane, D., & Hegarty, C. (2016). The Standardized Letter of Evaluation for Postgraduate Training. *Academic Medicine*, *91*(11), 1480-1482. http://dx.doi.org/10.1097/acm.0000000000001352

Maan, Z., Maan, I., Darzi, A., & Aggarwal, R. (2012). Systematic review of predictors of surgical performance. *British Journal of Surgery*, 99(12), 1610-1621. http://dx.doi.org/10.1002/bjs.8893

Makdisi, G., Takeuchi, T., Rodriguez, J., Rucinski, J., & Wise, L. (2011). How We Select Our Residents—A Survey of Selection Criteria in General Surgery Residents. *Journal of Surgical Education*, 68(1), 67-72. http://dx.doi.org/10.1016/j.jsurg.2010.10.003

Martin, J. (1996). Trainee selection for General Surgery. *ANZ Journal of Surgery*, 66(7), 428-430. http://dx.doi.org/10.1111/j.1445-2197.1996.tb00776.x

Martin, J., Blennerhassett, J., Hardman, D., & Mundy, J. (2009). Development of the surgical science examination of the Royal Australasian College of Surgeons surgical education and training programme: putting the chicken before the egg. *ANZ Journal of Surgery*, *79*(3), 169-174. http://dx.doi.org/10.1111/j.1445-2197.2008.04835.x

McManus, I. (2003). A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *BMJ*, 327(7407), 139-142. http://dx.doi.org/10.1136/bmj.327.7407.139

Medical Board of Australia - Registration. (2017). *Medicalboard.gov.au*. Retrieved 5 October 2017, from http://www.medicalboard.gov.au/Registration.aspx

Medical Education in Australia and New Zealand (pp. 1-4). Melbourne. Retrieved from https://www.surgeons.org/media/292226/medical_education_anz.pdf

Melendez, M., Xu, X., Sexton, T., Shapiro, M., & Mohan, E. (2008). The Importance of Basic Science and Clinical Research as a Selection Criterion for General Surgery Residency Programs. *Journal of Surgical Education*, *65*(2), 151-154. doi:10.1016/j.jsurg.2007.08.009

Michels, N., Avonts, M., Peeraer, G., Ulenaers, K., Van Gaal, L., & Bossaert, L. et al. (2016). Content validity of workplace-based portfolios: A multi-centre study. *Medical Teacher*, 1-10. http://dx.doi.org/10.3109/0142159x.2015.1132407

Miles, M., & Huberman, A. (1994). Qualitative data analysis: An expanded sourcebook (2nd ed., p. 18). Thousand Oaks: Sage Publications.

Miller, G. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65(9), S63-7. http://dx.doi.org/10.1097/00001888-199009000-00045

Minutes of a meeting of the Board of Surgical Education and Training held at RACS on Friday 8th October 2010. Melbourne. Unpublished

Monroe, A., Quinn, E., Samuelson, W., Dunleavy, D., & Dowd, K. (2013). An Overview of the Medical School Admission Process and Use of Applicant Data in Decision Making: What has changed since the 1980s?. *Academic Medicine*, *88*(5), 672-681. doi:10.1097/acm.0b013e31828bf252

Mukaka, M (2012), A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69–71.

Munro, D., Bore, M., & Powis, D. (2012). Pitfalls in using referee rating scales for postgraduate medical selection: they can be reliable but not valid. *Mededworld Publish*, 1(1), 1-19. Retrieved from http://mededworld.org/MedEdWorld-Papers/volume-1-2012.aspx

Murphy, D., Bruce, D., Mercer, S., & Eva, K. (2008). The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom. *Advances in Health Sciences Education*, 14(2), 219-232. http://dx.doi.org/10.1007/s10459-008-9104-8

National Ethnic Population Projections: 2013(base)–2038 (update). (2017). *Stats.govt.nz*. Retrieved 15 September 2017, from http://www.stats.govt.nz/browse_for_stats/population/estimates_and_projections/NationalEthnicPopulationProjections_HOTP2013-2038.aspx

Neely, D., Feinglass, J., & Wallace, W. (2010). Developing a Predictive Model to Assess Applicants to an Internal Medicine Residency. *Journal of Graduate Medical Education*, 2(1), 129-132. http://dx.doi.org/10.4300/jgme-d-09-00044.1

Newton, P. & Shaw, S. (2015). Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy and Practice*, 23(2), 178-197. http://dx.doi.org/10.1080/0969594x.2015.1037241

Nguyen, N., Biderman, M., & McDaniel, M. (2005). Effects of Response Instructions on Faking a Situational Judgment Test. *International Journal of Selection and Assessment*, *13*(4), 250-260. http://dx.doi.org/10.1111/j.1468-2389.2005.00322.x

Nicol, D. & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. http://dx.doi.org/10.1080/03075070600572090

Norcini, J. (2003). The Mini-CEX: A Method for Assessing Clinical Skills. *Annals of Internal Medicine*, *138*(6), 476. http://dx.doi.org/10.7326/0003-4819-138-6-200303180-00012

Norcini, J. (2005). The Mini Clinical Evaluation Exercise (mini-CEX). *The Clinical Teacher*, *2*(1), 25-30. http://dx.doi.org/10.1111/j.1743-498x.2005.00060.x

Norman, G. (2005). Editorial – Checklists vs. Ratings, the Illusion of Objectivity, the Demise of Skills and the Debasement of Evidence. *Advances in health sciences education: theory and practice 10*(1), 1-3. http://dx.doi.org/10.1007/s10459-005-4723-9

Nzags.co.nz,. (2016). *Become a Trainee » New Zealand Association of General Surgeons*. Retrieved 6 January 2016, from http://www.nzags.co.nz/index.php/education-and-training/become-a-trainee/

O'Brien, A., Harvey, J., Shannon, M., Lewis, K., & Valencia, O. (2011). A comparison of multiple mini-interviews and structured interviews in a UK setting. *Medical Teacher*, *33*(5), 397-402. http://dx.doi.org/10.3109/0142159x.2010.541532

Oldfield, Z., Beasley, S., Smith, J., Anthony, A., & Watt, A. (2013). Correlation of selection scores with subsequent assessment scores during surgical training. *ANZ Journal of Surgery*, 83(6), 412-416. http://dx.doi.org/10.1111/ans.12176

Pangaro, L., & ten Cate, O. (2013). Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Medical Teacher, 35*(6), e1197-e1210. doi:10.3109/0142159x.2013.788789

Parry, J., Mathers, J., Stevens, A., Parsons, A., Lilford, R., Spurgeon, P., & Thomas, H. (2006). Admissions processes for five year medical courses at English schools: review. *BMJ*, 332(7548), 1005-1009. http://dx.doi.org/10.1136/bmj.38768.590174.55

Patterson, F., Ashworth, V., Kerrin, M., & O'Neill, P. (2013). Situational judgement tests represent a measurement method and can be designed to minimise coaching effects. *Medical Education*, *47*(2), 220-221. http://dx.doi.org/10.1111/medu.12100

Patterson, F., Baron, H., Carr, V., Plint, S., & Lane, P. (2009). Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. *Medical Education*, *43*(1), 50-57. doi:10.1111/j.1365-2923.2008.03238.x

Patterson, F., Ferguson, E., Norfolk, T., & Lane, P. (2005). A new selection system to recruit general practice registrars: preliminary findings from a validation study. *BMJ*, *330*(7493), 711-714. doi:10.1136/bmj.330.7493.711

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education*, 50: 36-60. doi: 10.1111/medu.12817

Patterson, F., Lievens, F., Kerrin, M., Munro, N., & Irish, B. (2013). The predictive validity of selection for entry into postgraduate training in general practice: evidence from three longitudinal studies. *British Journal of General Practice*, 63(616), 734-741. http://dx.doi.org/10.3399/bjgp13x674413

Patterson, F., Rowett, E., Hale, R., Grant, M., Roberts, C., Cousans, F., & Martin, S. (2016). The predictive validity of a situational judgement test and multiple-mini interview for entry into postgraduate training in Australia. *BMC Medical Education*, *16*(1). http://dx.doi.org/10.1186/s12909-016-0606-4

Patterson, F., Zibarras, L., Carr, V., Irish, B., & Gregory, S. (2011). Evaluating candidate reactions to selection practices using organisational justice theory. *Medical Education*, *45*(3), 289-297. http://dx.doi.org/10.1111/j.1365-2923.2010.03808.x

Pau, A., Jeevaratnam, K., Chen, Y., Fall, A., Khoo, C., & Nadarajah, V. (2013). The Multiple Mini-Interview (MMI) for student selection in health professions training – A systematic review. *Medical Teacher*, *35*(12), 1027-1041. http://dx.doi.org/10.3109/0142159x.2013.829912

Pearson Product-Moment Correlation - When you should run this test, the range of values the coefficient can take and how to measure strength of association. (2017). *Statistics.laerd.com*. Retrieved 25 May 2017, from https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php

Pollett, W., & Waxman, B. (2012). Postgraduate surgical education and training in Canada and Australia: each may benefit from the other's experiences. *ANZ Journal of Surgery*, *82*(9), 581-587. http://dx.doi.org/10.1111/j.1445-2197.2012.06172.x

Ponnamperuma, G. (2010). Selection for postgraduate training. *The Clinical Teacher*, *7*(4), 276-280. doi:10.1111/j.1743-498x.2010.00400.x

Poole, P., Shulruf, B., Rudland, J., & Wilkinson, T. (2012). Comparison of UMAT scores and GPA in prediction of performance in medical school: a national study. *Medical Education*, *46*(2), 163-171. http://dx.doi.org/10.1111/j.1365-2923.2011.04078.x

Population clock. (2015). *Stats.govt.nz*. Retrieved 22 November 2015, from http://www.stats.govt.nz/tools_and_services/population_clock.aspx

Prager, J., Myer III, C., & Pensak, M. (2010). Improving the letter of recommendation. *Otolaryngology - Head and Neck Surgery*, 143(3), 327-330. http://dx.doi.org/10.1016/j.otohns.2010.03.017

Prager, J., Myer, C., Hayes, K., Myer, C., & Pensak, M. (2010). Improving Methods of Resident Selection. *The Laryngoscope*, *120*(S3), S20-S20. http://dx.doi.org/10.1002/lary.21198

Prideaux, D., Roberts, C., Eva, K., Centeno, A., McCrorie, P., & McManus, C. et al. (2011). Assessment for selection for the health care professions and specialty training: Consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher,* 33(3), 215-223. http://dx.doi.org/10.3109/0142159x.2011.551560

Principles for the Validation and use of Personnel Selection Procedures. (2003) (4th ed.). Bowling Green, Ohio. Retrieved from http://www.siop.org/_principles/principles.pdf

Razack, S. (2016). 'Fairness' and student selection: the case for mandatory air quotes. *Medical Education*, *50*(6), 600-602. http://dx.doi.org/10.1111/medu.12997

Razack, S., Hodges, B., Steinert, Y., & Maguire, M. (2014). Seeking inclusion in an exclusive process: discourses of medical school student selection. *Medical Education*, *49*(1), 36-47. http://dx.doi.org/10.1111/medu.12547

Rcsi.ie,. (2015). *Applying to ST1 - ST2 - Royal College Surgeons in Ireland*. Retrieved 2 October 2015, from http://www.rcsi.ie/index.jsp?p=1609&n=1854&a=0

Reid, C., Kim, D., Mandel, J., Smith, A., & Bansal, V. (2014). Correlating surgical clerkship evaluations with performance on the National Board of Medical Examiners examination. *Journal of Surgical Research*, 190(1), 29-35. http://dx.doi.org/10.1016/j.jss.2014.02.031

Reiter, H. & McConnell, M. (2014). How I learned to stop worrying and love the bias. *Medical Education*, 48(12), 1141-1143. http://dx.doi.org/10.1111/medu.12602

Reiter, H., Eva, K., Rosenfeld, J., & Norman, G. (2007). Multiple mini-interviews predict clerkship and licensing examination performance. *Medical Education*, *41*(4), 378-384. http://dx.doi.org/10.1111/j.1365-2929.2007.02709.x

Robb, N., Dunkley, L., Boynton, P., & Greenhalgh, T. (2007). Looking for a better future: Identity construction in socio-economically deprived 16-year olds considering a career in medicine. *Social Science and Medicine*, *65*(4), 738-754. http://dx.doi.org/10.1016/j.socscimed.2007.03.011

Roberts, C., & Togno, J. (2011). Selection into specialist training programs: an approach from general practice. *Medical Journal of Australia*, *194*(2), 93-95.

Roberts, C., Clark, T., Burgess, A., Frommer, M., Grant, M., & Mossman, K. (2014). The validity of a behavioural multiple-mini-interview within an assessment centre for selection into specialty training. *BMC Medical Education, 14*(1), 169-179. http://dx.doi.org/10.1186/1472-6920-14-169

Roberts, C., Khanna, P., Rigby, L., Bartle, E., Llewellyn, A., & Gustavs, J. et al. (2017). Utility of selection methods for specialist medical training: A BEME (best evidence medical education) systematic review: BEME guide no. 45. *Medical Teacher*, 3-19. http://dx.doi.org/10.1080/0142159x.2017.1367375

Roberts, C., Walton, M., Rothnie, I., Crossley, J., Lyon, P., Kumar, K., & Tiller, D. (2008). Factors affecting the utility of the multiple mini-interview in selecting candidates for graduate-entry medical school. *Medical Education*, 42(4), 396-404. http://dx.doi.org/10.1111/j.1365-2923.2008.03018.x

Robins, J., McInnes, M., & Esmail, K. (2014). What information is provided in transcripts and Medical Student Performance Records from Canadian Medical Schools? A retrospective cohort study. *Medical Education Online*, *19*(1). http://dx.doi.org/10.3402/meo.v19.25181

Rockstuhl, T., Ang, S., Ng, K., Lievens, F., & Van Dyne, L. (2015). Putting judging situations into situational judgment tests: Evidence from intercultural multimedia SJTs. *Journal of Applied Psychology*, *100*(2), 464-480. http://dx.doi.org/10.1037/a0038098

Rosenfeld, J., Reiter, H., Trinh, K., & Eva, K. (2008). A Cost Efficiency Comparison Between the Multiple Mini-Interview and Traditional Admissions Interviews. *Advances in Health Sciences Education*, *13*(1), 43-58. http://dx.doi.org/10.1007/s10459-006-9029-z

Rousseau, D., & Greller, M. (1994). Human resource practices: Administrative contract makers. *Human Resource Management*, *33*(3), 385-401. doi:10.1002/hrm.3930330308

Royal Australasian College of Surgeons Accreditation Submission 2016. (2016) (pp. 1-146). Melbourne. Retrieved from https://surgeons.org/media/24925838/2016-12-21_rpt_amc_accreditation.pdf

Royal Australasian College of Surgeons Diversity and Inclusion Plan. (2017) (1st ed., pp. 1-15). Melbourne. Retrieved from https://surgeons.org/media/24924140/2016_12_20_diversity_and_inclusion_plan.pdf

Royal Australasian College of Surgeons, 2007 Executive Summary. (2007) (pp. 1-6). Canberra. Retrieved from https://www.amc.org.au/joomla-files/images/SpecialistCollegeReports/RACS.pdf

Royal College of Physicians and Surgeons of Canada: CanMEDS Framework. (2016). *Royalcollege.ca*. Retrieved 10 April 2016, from http://www.royalcollege.ca/portal/page/portal/rc/canmeds/framework.

Rutkow, I. (2013). The Education, Training, and Specialization of Surgeons. *Annals of Surgery*, *258*(6), 1130-1136. doi:10.1097/sla.0b013e3182a6f6a6

Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in Health Sciences Education*, 6, 159-175.

Sandow, P., Jones, A., Peek, C., Courts, F., & Watson, R. (2002). Correlation of admission criteria with dental school performance and attrition. *Journal of Dental Education*, *66*(3), 385-392.

Scanlon Foundation, (2016). Multiculturalism Discussion Paper. Melbourne: Scanlon Foundation. Retrieved from http://scanlonfoundation.org.au/wp-content/uploads/2016/02/DiscussionPaperMulticulturalismFINAL.pdf

Schaufeli WB, Buunk BP. Professional burnout. In: Schabracq MJ, Winnubst JAM, Cooper CL, editors. *Handbook of Work and Health Psychology*. New York: Wiley; 1996. pp. 311–346.

Schaverien, M. (2016). Selection for Surgical Training: An Evidence-Based Review. *Journal of Surgical Education*, *73*(4), 721-729. http://dx.doi.org/10.1016/j.jsurg.2016.02.007

Schripsema, N., van Trigt, A., Borleffs, J., & Cohen-Schotanus, J. (2014). Selection and study performance: comparing three admission processes within one medical school. *Medical Education, 48*(12), 1201-1210. doi:10.1111/medu.12537

Schueneman, A., Pickleman, J., Hesslein, R., & Freeark, R. (1984). Neuropsychologic predictors of operative skill among general surgery residents. *Surgery*, 96(2), 288-295.

Schuwirth, L. & van der Vleuten, C. (2011). Programmatic assessment: From assessment of learning to assessment for learning. *Medical Teacher*, 33(6), 478-485. http://dx.doi.org/10.3109/0142159x.2011.565828

Schuwirth, L. & van der Vleuten, C. (2012). Programmatic assessment and Kane's validity perspective. *Medical Education*, 46, 38-48. doi: 10.1111/j.1365-2923.2011.04098.x

Seabott, H., Smith, R., Alseidi, A., & Thirlby, R. (2012). The Surgical Residency Interview: A Candidate-Centered, Working Approach. *Journal of Surgical Education*, *69*(6), 802-806. http://dx.doi.org/10.1016/j.jsurg.2012.06.004

Searle, J., & McHarg, J. (2003). Selection for medical school: just pick the right students and the rest is easy!. *Medical Education, 37*(5), 458-463. doi:10.1046/j.1365-2923.2003.01496.x

Sharma, N. (2015). Medical students' perceptions of the situational judgement test: a mixed methods study. *British Journal of Hospital Medicine*, 76(4), 234-238. http://dx.doi.org/10.12968/hmed.2015.76.4.234

Shellito, J., Osland, J., Helmer, S., & Chang, F. (2010). American Board of Surgery examinations: can we identify surgery residency applicants and residents who will pass the examinations on the first attempt?. *The American Journal of Surgery*, 199(2), 216-222. http://dx.doi.org/10.1016/j.amjsurg.2009.03.006

Shortell, T. (2017). *Sociology 712*. *Shortell.org*. Retrieved 25 May 2017, from http://www.shortell.org/book/chap18.html

Sims, R. (1994). Human resource management's role in clarifying the new psychological contract. *Human Resource Management*, *33*(3), 373-382. doi:10.1002/hrm.3930330306

Singh, P., Aggarwal, R., Pucher, P., Duisberg, A., Arora, S., & Darzi, A. (2014). Defining quality in surgical training: perceptions of the profession. *The American Journal of Surgery*, 207(4), 628-636. http://dx.doi.org/10.1016/j.amjsurg.2013.07.044

Singletary, S. (2010). A Fire in Our Hearts: Passion and the Art of Surgery. *Annals of Surgical Oncology*, *17*(2), 364-370. doi:10.1245/s10434-009-0732-x

Siu, E., & Reiter, H. (2009). Overview: what's worked and what hasn't as a guide towards predictive admissions tool development. *Advances in Health Sciences Education*, *14*(5), 759-775. http://dx.doi.org/10.1007/s10459-009-9160-8

Sladek, R., Bond, M., Frost, L., & Prior, K. (2016). Predicting success in medical school: a longitudinal study of common Australian student selection tools. *BMC Medical Education*, *16*(1). http://dx.doi.org/10.1186/s12909-016-0692-3

Specialist Medical Colleges. (2017) (pp. 1-3). Canberra. Retrieved from https://www.amc.org.au/files/bf4d51290e53ba5d685f0e491014f45156000ba0_original.pdf

Stemig, M., Sackett, P., & Lievens, F. (2015). Effects of Organizationally Endorsed Coaching on Performance and Validity of Situational Judgment Tests. *International Journal of Selection and Assessment*, *23*(2), 174-181. http://dx.doi.org/10.1111/ijsa.12105

Storey, C. (2014). A brief history of the specialties from Federation to the present. *The Medical Journal of Australia*, *201*(1), 26-28. doi:10.5694/mja14.00136

Surgeons.org,. (2016). *Board of SET | Royal Australasian College of Surgeons*. Retrieved 6 January 2016, from http://www.surgeons.org/about/governance-committees/committees/education-board/set-board/

Surgicalcareers.rcseng.ac.uk,. (2015). *Application Form — Surgical Careers | The Royal College of Surgeons of England*. Retrieved 2 October 2015, from http://surgicalcareers.rcseng.ac.uk/students/school-student/applying-to-medical-school/application-form

Syme, G. (1928). The aims and objects of the College of Surgeons of Australasia. *The Medical Journal of Australia*, (April), 488-491.

Tang, C., Hilsinger, R., Cruz, R., Schloegel, L., Byl, F., & Rasgon, B. (2014). Manual Dexterity Aptitude Testing. *JAMA Otolaryngology Head and Neck Surgery*, *140*(3), 243-249. http://dx.doi.org/10.1001/jamaoto.2013.6456

Tansley, P., Kakar, S., Withey, S., & Butler, P. (2007). Visuospatial and Technical Ability in the Selection and Assessment of Higher Surgical Trainees in the London Deanery. *The Annals of The Royal College of Surgeons of England*, *89*(6), 591-595. http://dx.doi.org/10.1308/003588407x187702

Tavakol, M. & Dennick, R. (2011). Post-examination analysis of objective tests. *Medical Teacher*, 33(6), 447-458. http://dx.doi.org/10.3109/0142159x.2011.564682

Taylor, I. (2005). Can performance as an undergraduate assist entry selection into surgical training programmes?. *Annals of The Royal College of Surgeons of England*, *87*(1), 1-2. doi:10.1308/1478708051612

ten Cate, O. (2007). Medical education in the Netherlands. *Medical Teacher, 29*(8), 752-757. doi:10.1080/01421590701724741 ten Cate, O., Tobin, S., & Stokes, M. (2017). Bringing competencies closer to day-to-day clinical work through entrustable professional activities. *The Medical Journal of Australia*, *206*(1), 14-16. http://dx.doi.org/10.5694/mja16.00481

The Panel on Fair Access to the Professions. (2009). *Unleashing Aspiration: The Final Report of the Panel on Fair Access to the Professions*. London: Cabinet Office. Retrieved from http://www.cabinetoffice.gov.uk/accessprofessions

The trainee doctor. (2011) (pp. 1-58). London. Retrieved from http://www.gmc-uk.org/Trainee_Doctor.pdf_39274940.pdf

Thomas, H., Davison, I., Gee, H., Grant, J., & Taylor, C. (2013). The fairness, effectiveness and acceptability of selection for specialty training in the UK. *British Journal of Hospital Medicine*, *74*(1), 47-51. http://dx.doi.org/10.12968/hmed.2013.74.1.47

Thordarson, D., Ebramzadeh, E., Sangiorgio, S., Schnall, S., & Patzakis, M. (2007). Resident Selection: How are we doing and why?. *Clinical Orthopaedics and Related Research*, 459, 255-259. http://dx.doi.org/10.1097/BLO.0b013e31805d7eda

Tooke, J. (2008). *Aspiring to excellence: Final report of the independent inquiry into modernising medical careers*. London: Aldridge Press.

Tsouroufli, M., & Malcolm, I. (2014). Equality, diversity and fairness in medical education: international perspectives. *Medical Education 49*(1), 4-6. http://dx.doi.org/10.1111/medu.12601

Twycross, A. & Shields, L. (2004). Validity and reliability–What's it all about? Part 1 Validity in quantitative studies. *Paediatric Care*, 16(9), 28-28. http://dx.doi.org/10.7748/paed2004.11.16.9.28.c954

Urlings-Strop, L., Themmen, A., Stijnen, T., & Splinter, T. (2011). Selected medical students achieve better than lottery-admitted students during clerkships. *Medical Education*, *45*(10), 1032-1040. doi:10.1111/j.1365-2923.2011.04031.x

Van Bruwaene, S., Lissens, A., De Win, G., Neyrinck, B., Lens, W., Schijven, M., & Miserez, M. (2015). Surgical Skill: Trick or Trait?. *Journal of Surgical Education*, 72(6), 1247-1253. http://dx.doi.org/10.1016/j.jsurg.2015.05.004

van de Loo, R. (1988). Selection of surgical trainees in The Netherlands. *Annals of The Royal College of Surgeons of England*, 70, 277-9.

van den Bos, K. and Miedema, J. (2000). Toward understanding why fairness matters: the influence of mortality salience on reactions to procedural fairness. *Journal of Personality and Social Psychology*, 79(3), 355.

van den Bos, K., Wilke, H., & Lind, E. (1998). When do we need procedural fairness? The role of trust in authority. *Journal of Personality and Social Psychology*, *75*(6), 1449-1458.

van der Vleuten, C. (2000). Validity of final examinations in undergraduate medical training. *BMJ*, *321*(7270), 1217-1219. doi:10.1136/bmj.321.7270.1217

van der Vleuten, C., Schuwirth, L., Scheele, F., Driessen, E., & Hodges, B. (2010). The assessment of professional competence: building blocks for theory development. *Best Practice and Research Clinical Obstetrics and Gynaecology*, 24(6), 703-719. http://dx.doi.org/10.1016/j.bpobgyn.2010.04.001

Vassiliou, M., & Feldman, L. (2011). Objective assessment, selection, and certification in surgery. *Surgical Oncology*, *20*(3), 140-145. doi:10.1016/j.suronc.2010.09.004

von Websky, M., Oberkofler, C., Rufibach, K., Raptis, D., Lehmann, K., Hahnloser, D., & Clavien, P. (2012). Trainee satisfaction in surgery residency programs: Modern management tools ensure trainee motivation and success. *Surgery*, *152*(5), 794-801. doi:10.1016/j.surg.2011.07.037

Walker, I. (2008). *Reporting results*. *Staff.bath.ac.uk*. Retrieved 8 October 2017, from http://staff.bath.ac.uk/pssiw/stats2/page2/page3/page3.html

Wanzel, K., Hamstra, S., Caminiti, M., Anastakis, D., Grober, E., & Reznick, R. (2003). Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. *Surgery*, 134(5), 750-757. http://dx.doi.org/10.1016/s0039-6060(03)00248-4

Wanzel, K., Ward, M., & Reznick, R. (2002). Teaching the surgical craft: From selection to certification. *Current Problems in Surgery*, *39*(6), 583-660. doi:10.1067/mog.2002.123481

Warren, M. (1951). Medical education during the eighteenth century. *Postgraduate Medical Journal*, *27*, 304-311. doi:10.1136/pgmj.27.308.304

Watters, D., & Civil, I. (2011). *Surgical competence and performance guide*. Melbourne: Royal Australasian College of Surgeons. Retrieved from http://www.surgeons.org/media/18955288/surgical_competence_and_performance_guide__2011_.pdf

White, A. (2002). Resident selection: Are we putting the cart before the horse?. *Clinical Orthopaedics and Related Research*, 399, 255-259. http://dx.doi.org/10.1097/00003086-200206000-00032

White, J., Brownell, K., Lemay, J., & Lockyer, J. (2012). "What Do They Want Me To Say?" The hidden curriculum at work in the medical school selection process: a qualitative study. *BMC Med Educ*, *12*(1), 17. http://dx.doi.org/10.1186/1472-6920-12-17

Wiliam, D. (2001). Reliability, validity, and all that jazz. *Education* 3-13, 29(3), 17-21. http://dx.doi.org/10.1080/03004270185200311

Wilkinson, D., Zhang, J., Byrne, G., Luke, H., Ozolins, I., Parker, M., & Peterson, R. (2008). Medical school selection criteria and the prediction of academic performance. *The Medical Journal of Australia*, *188*(6), 349–354.

Williams, R., Klamen, D., & McGaghie, W. (2003). SPECIAL ARTICLE: Cognitive, Social and Environmental Sources of Bias in Clinical Performance Ratings. *Teaching and Learning in Medicine*, 15(4), 270-292. http://dx.doi.org/10.1207/s15328015tlm150411

Williams, R., Verhulst, S., Colliver, J., Sanfey, H., Chen, X., & Dunnington, G. (2012). A template for reliable assessment of resident operative performance: Assessment intervals, numbers of cases and raters. *Surgery*, 152(4), 517-527. http://dx.doi.org/10.1016/j.surg.2012.07.004

Workplace Based Assessment: A guide for implementation. (2010) (pp. 1-12). London. Retrieved from http://www.gmc-uk.org/Workplace_Based_Assessment___A_guide_for_implementation_0410.pdf_48905168.pdf

Yeates, P., O'Neill, P., Mann, K., & Eva, K. (2012). Seeing the same thing differently. *Advances in Health Sciences Education*, 18(3), 325-341. http://dx.doi.org/10.1007/s10459-012-9372-1

Yoshimura, H., Kitazono, H., Fujitani, S., Machi, J., Saiki, T., Suzuki, Y., & Ponnamperuma, G. (2015). Past-behavioural versus situational questions in a postgraduate admissions multiple mini-interview: a reliability and acceptability comparison. *BMC Medical Education*, *15*(1). http://dx.doi.org/10.1186/s12909-015-0361-y

Young, T. (1997). Teaching medical students to lie. The disturbing contradiction: medical ideals and the resident-selection process. *Canadian Medical Association Journal*, 156(2), 219-22.

APPENDICES

APPENDIX A – Supplementary tables and figures

Candidate performance in selection, examinations, DOPS, MiniCEX and ETAs.

Descriptive statistics: Tables

Table 37 *General Surgery candidate performance in selection items*

		CV scores		Referee Report scores		Interview scores		Total selection scores	
Cohort	**n**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
2008 Au	81	51.65	13.80	76.92	9.15	80.92	11.14	73.47	6.38
2008 NZ	19	52.63	15.20	76.33	7.23	80.30	5.39	73.18	4.45
2008 ANZ	100	51.84	14.00	76.81	8.79	80.80	10.28	73.41	6.05
2009 Au	89	42.52	13.70	80.43	7.27	83.63	6.86	74.13	3.75
2009 NZ	18	46.67	11.74	80.76	5.69	80.83	5.97	73.97	4.26
2009 ANZ	107	43.22	13.43	80.48	7.01	83.16	6.78	74.10	3.81
2010 Au	121	52.10	13.71	84.91	6.82	82.34	5.99	77.37	3.61
2010 NZ	19	54.00	8.92	88.24	4.39	73.16	6.12	75.36	3.83
2010 ANZ	140	52.36	13.15	85.37	6.63	81.10	6.77	77.10	3.69
08,09,10 Au	291	49.04	14.36	81.32	8.33	82.34	8.03	75.29	4.90
08,09,10 NZ	56	51.18	12.43	80.76	5.69	78.05	6.74	73.97	4.26
08,09,10 ANZ	347	49.39	14.07	81.40	8.22	81.65	7.98	75.11	4.81
Summary									
Min	18	42.52	8.92	76.33	4.39	73.16	5.39	73.18	3.61
Max	347	54.00	15.20	88.24	9.15	83.63	11.14	77.37	6.38
Mean per year	58	49.93	12.85	81.27	6.76	80.20	6.91	74.58	4.38

Note: Mean per year does not include 'Combined years' or ANZ

Table 38 *General Surgery trainee examination performance*

	Generic SSE score			Specialty SSE score			Clinical Examination score		
Cohort	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
2008 Au	73.57	5.48	78	73.46	5.57	56	65.76	6.70	79
2008 NZ	73.43	5.31	18	71.76	4.68	10	64.42	5.61	18
2008 ANZ	73.54	5.42	96	73.20	5.44	66	65.51	6.51	97
2009 Au	73.07	5.82	88	73.15	5.75	80	68.40	5.98	88
2009 NZ	75.07	5.21	17	72.88	4.06	13	69.52	6.27	17
2009 ANZ	73.39	5.75	105	73.12	5.53	93	68.58	6.01	105
2010 Au	72.11	6.62	118	71.86	7.40	114	68.90	5.15	116
2010 NZ	75.86	5.77	19	76.06	7.77	18	68.12	6.14	19
2010 ANZ	72.63	6.62	137	72.43	7.56	132	68.79	5.28	135
08,09,10 Au	72.81	6.09	284	72.63	6.53	250	67.87	6.01	283
08,09,10 NZ	74.80	5.44	54	74.00	6.25	41	67.33	6.28	54
08,09,10 ANZ	73.13	6.03	338	72.82	6.50	291	67.78	6.04	337
Summary									
Min	72.11	5.21	17	71.76	4.06	10	64.42	5.15	17
Max	75.86	6.62	338	76.06	7.77	291	69.52	6.70	337
Mean per year	73.85	5.70	56	73.20	5.87	49	67.52	5.98	56

Note: Mean per year does not include 'Combined years' or ANZ

Table 39 *General Surgery trainee DOPS performance, n ≥ 5*

	DOPS 1			DOPS 2			DOPS 3			DOPS 4			Average DOPS		
Cohort	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
2008 Au	79.78	12.16	45	78.59	10.88	39	85.78	14.18	20	84.16	11.79	10	79.71	8.98	45
2008 NZ	84.76	10.84	14	85.37	11.06	11	Y	Y	2	Q	Q	0	85.38	8.73	14
2008 ANZ	80.96	11.96	59	80.08	11.17	50	85.24	13.88	22	84.16	11.79	10	81.06	9.18	59
2009 Au	79.89	12.44	64	83.06	12.64	61	81.79	11.17	30	81.37	13.31	13	80.95	9.62	66
2009 NZ	79.59	12.52	10	85.42	14.13	8	Y	Y	2	Y	Y	2	82.42	8.62	10
2009 ANZ	79.85	12.36	74	83.34	12.74	69	82.51	11.29	32	80.07	12.81	15	81.14	9.46	76
2010 Au	81.01	12.08	102	80.72	12.83	77	81.42	13.63	34	81.84	12.97	17	81.42	9.91	103
2010 NZ	78.76	9.17	17	80.09	9.85	8	Q	Q	0	Q	Q	0	79.87	7.16	17
2010 ANZ	80.69	11.70	119	80.66	12.54	85	81.42	13.63	34	81.84	12.97	17	81.20	9.55	120
08,09,10 Au	80.41	12.16	211	81.06	12.41	177	82.59	12.91	84	82.27	12.53	40	80.91	9.61	214
08,09,10 NZ	81.01	10.70	41	83.82	11.54	27	Y	Y	4	Y	Y	2	82.37	8.24	41
08,09,10 ANZ	80.51	11.92	252	81.42	12.31	204	82.77	12.83	88	81.76	12.43	42	81.15	9.40	255
Summary															
Min	78.76	9.17	10	78.59	9.85	8	81.42	11.17	0	80.07	11.79	0	79.71	7.16	10
Max	84.76	12.52	252	85.42	14.13	204	85.78	14.18	88	84.16	13.31	42	85.38	9.91	255
Mean per year	80.63	11.54		82.21	11.90		83.00	12.99		82.46	12.69		81.63	8.84	

Note 1: Y indicates no scores listed because n < 5 (see Table 42 General Surgery trainee DOPS performance, including n < 5)

Note 2: Q indicates no scores listed because n = 0 (see Table 42 General Surgery trainee DOPS performance, including n < 5)

Note 3: Mean per year does not include 'Combined years' or ANZ

Table 40 *General Surgery trainee MiniCEX performance, n ≥ 5*

	MiniCEX 1			MiniCEX 2			MiniCEX 3			MiniCEX 4			Average MiniCEX		
Cohort	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
2008 Au	84.91	11.92	46	85.99	12.52	41	89.93	12.03	20	79.72	20.02	9	85.15	8.62	47
2008 NZ	84.39	11.22	14	86.52	11.92	10	Y	Y	2	Q	Q	0	86.07	9.49	14
2008 ANZ	84.79	11.67	60	86.09	12.29	51	89.52	12.37	22	79.72	20.02	9	85.36	8.75	61
2009 Au	82.76	13.91	67	86.73	11.33	58	83.18	11.70	24	86.85	13.42	10	83.53	11.14	68
2009 NZ	87.41	11.61	10	91.26	10.49	8	Y	Y	2	Y	Y	2	89.93	7.17	10
2009 ANZ	83.36	13.65	77	87.28	11.25	66	84.19	11.82	26	87.50	12.26	12	84.35	10.89	78
2010 Au	83.34	12.41	100	82.41	14.85	74	84.71	11.48	34	86.31	12.39	12	83.18	10.48	101
2010 NZ	81.47	8.52	17	91.64	9.08	8	Q	Q	0	Q	Q	0	84.20	7.66	17
2010 ANZ	83.06	11.91	117	83.31	14.61	82	84.71	11.48	34	86.31	12.39	12	83.32	10.11	118
08,09,10 Au	83.49	12.76	213	84.71	13.29	173	85.58	11.83	78	84.57	15.08	31	83.72	10.31	216
08,09,10 NZ	83.92	10.29	41	89.55	10.54	26	Y	Y	4	Y	Y	2	86.23	8.34	41
08,09,10 ANZ	83.56	12.38	254	85.34	13.04	199	85.83	11.90	82	84.95	14.68	33	84.12	10.05	257
Summary															
Min	81.47	8.52	10	82.41	9.08	8	83.18	11.48	0	79.72	12.26	0	83.18	7.17	10
Max	87.41	13.91	254	91.64	14.85	199	89.93	12.37	82	87.5	20.02	33	89.93	11.14	257
Mean per year	84.05	11.60		87.43	11.70		85.94	11.74		84.29	15.28		85.34	9.09	

Note 1: Y indicates no scores listed because n < 5 (see Table 43 General Surgery trainee MiniCEX performance, including n < 5)
Note 2: Q indicates no scores listed because n = 0 (see Table 43 General Surgery trainee MiniCEX performance, including n < 5
Note 3: Mean per year does not include 'Combined years' or ANZ

Table 41 *General Surgery trainee End of Term Assessment performance, n ≥ 5*

	ETA 1%			ETA 2%			ETA 3%			ETA 4%			Average ETA%		
Cohort	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
2008 Au	84.50	6.42	59	85.12	7.01	60	85.30	6.92	55	85.17	6.79	50	84.39	4.80	62
2008 NZ	84.71	5.73	19	87.73	5.14	19	84.14	5.62	19	84.98	5.78	19	85.39	3.92	19
2008 ANZ	84.55	6.22	78	85.75	6.67	79	85.00	6.60	74	85.11	6.48	69	84.63	4.60	81
2009 Au	85.04	6.61	77	84.06	5.55	75	86.07	7.66	68	83.96	6.78	56	84.82	4.44	79
2009 NZ	85.52	7.56	16	83.89	5.63	15	79.30	6.37	7	83.18	4.79	14	83.82	5.18	18
2009 ANZ	85.12	6.74	93	84.03	5.53	90	85.44	7.77	75	83.81	6.41	70	84.63	4.58	97
2010 Au	82.66	6.24	90	83.74	5.88	110	Q	Q	0	Q	Q	0	83.21	4.93	116
2010 NZ	84.05	6.27	16	86.54	6.81	18	Q	Q	0	Q	Q	0	85.71	5.79	18
2010 ANZ	82.87	6.24	106	84.14	6.07	128	Q	Q	0	Q	Q	0	83.54	5.10	134
08,09,10 Au	83.95	6.48	226	84.18	6.08	245	85.73	7.32	123	84.53	6.78	106	83.99	4.79	257
08,09,10 NZ	84.76	6.41	51	86.21	6.00	52	82.84	6.11	26	84.22	5.38	33	84.98	4.98	55
08,09,10 ANZ	84.10	6.46	277	84.53	6.11	297	85.23	7.19	149	84.46	6.46	139	84.16	4.83	312
Summary															
Min	82.66	5.73	16	83.74	5.14	15	79.30	5.62	0	83.18	4.79	0	83.21	3.92	18
Max	85.52	7.56	277	87.73	7.01	297	86.07	7.66	149	85.17	6.79	139	85.71	5.79	312
Mean per year	84.41	6.47		85.18	6.00		83.70	6.64		84.32	6.04		84.56	4.84	

Note 1: Q indicates no scores listed because n = 0
Note 2: Mean per year does not include 'Combined years' or ANZ

Table 42 *General Surgery trainee DOPS performance, including n < 5*

	DOPS 1			DOPS 2			DOPS 3			DOPS 4			Average DOPS		
Cohort	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
2008 Au	79.78	12.16	45	78.59	10.88	39	85.78	14.18	20	84.16	11.79	10	79.71	8.98	45
2008 NZ	84.76	10.84	14	85.37	11.06	11	79.86	12.77	2	-	-	0	85.38	8.73	14
2008 ANZ	80.96	11.96	59	80.08	11.17	50	85.24	13.88	22	84.16	11.79	10	81.06	9.18	59
2009 Au	79.89	12.44	64	83.06	12.64	61	81.79	11.17	30	81.37	13.31	13	80.95	9.62	66
2009 NZ	79.59	12.52	10	85.42	14.13	8	93.34	9.43	2	71.67	2.35	2	82.42	8.62	10
2009 ANZ	79.85	12.36	74	83.34	12.74	69	82.51	11.29	32	80.07	12.81	15	81.14	9.46	76
2010 Au	81.01	12.08	102	80.72	12.83	77	81.42	13.63	34	81.84	12.97	17	81.42	9.91	103
2010 NZ	78.76	9.17	17	80.09	9.85	8	-	-	0	-	-	0	79.87	7.16	17
2010 ANZ	80.69	11.70	119	80.66	12.54	85	81.42	13.63	34	81.84	12.97	17	81.20	9.55	120
08,09,10 Au	80.41	12.16	211	81.06	12.41	177	82.59	12.91	84	82.27	12.53	40	80.91	9.61	214
08,09,10 NZ	81.01	10.70	41	83.82	11.54	27	86.60	12.02	4	71.67	2.35	2	82.37	8.24	41
08,09,10 ANZ	80.51	11.92	252	81.42	12.31	204	82.77	12.83	88	81.76	12.43	42	81.15	9.40	255
Summary															
Min	78.76	9.17	10	78.59	9.85	8	79.86	9.43	2	71.67	2.35	2	79.71	7.16	10
Max	84.76	12.52	252	85.42	14.13	204	93.34	14.18	88	84.16	13.31	42	85.38	9.91	255
Mean per year	80.63	11.53		82.21	11.90		84.44	12.23		79.76	10.11		81.62	8.84	

Note 1: Mean per year does not include 'Combined years' or ANZ

Table 43 *General Surgery trainee MiniCEX performance, including n < 5*

	MiniCEX 1			MiniCEX 2			MiniCEX 3			MiniCEX 4			Average MiniCEX		
Cohort	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
2008 Au	84.91	11.92	46	85.99	12.52	41	89.93	12.03	20	79.72	20.02	9	85.15	8.62	47
2008 NZ	84.39	11.22	14	86.52	11.92	10	85.42	20.63	2	-	-	0	86.07	9.49	14
2008 ANZ	84.79	11.67	60	86.09	12.29	51	89.52	12.37	22	79.72	20.02	9	85.36	8.75	61
2009 Au	82.76	13.91	67	86.73	11.33	58	83.18	11.70	24	86.85	13.42	10	83.53	11.14	68
2009 NZ	87.41	11.61	10	91.26	10.49	8	96.30	5.24	2	90.74	2.62	2	89.93	7.17	10
2009 ANZ	83.36	13.65	77	87.28	11.25	66	84.19	11.82	26	87.50	12.26	12	84.35	10.89	78
2010 Au	83.34	12.41	100	82.41	14.85	74	84.71	11.48	34	86.31	12.39	12	83.18	10.48	101
2010 NZ	81.47	8.52	17	91.64	9.08	8	-	-	0	-	-	0	84.20	7.66	17
2010 ANZ	83.06	11.91	117	83.31	14.61	82	84.71	11.48	34	86.31	12.39	12	83.32	10.11	118
08,09,10 Au	83.49	12.76	213	84.71	13.29	173	85.58	11.83	78	84.57	15.08	31	83.72	10.31	216
08,09,10 NZ	83.92	10.29	41	89.55	10.54	26	90.86	13.80	4	90.74	2.62	2	86.23	8.34	41
08,09,10 ANZ	83.56	12.38	254	85.34	13.04	199	85.83	11.90	82	84.95	14.68	33	84.12	10.05	257
Summary															
Min	81.47	8.52	10	82.41	9.08	8	83.18	5.24	2	79.72	2.62	2	83.18	7.17	10
Max	87.41	13.91	254	91.64	14.85	199	96.30	20.63	82	90.74	20.02	33	89.93	11.14	257
Mean	84.05	11.60		87.42	11.70		87.91	12.21		85.91	12.11		85.34	9.10	

Note 1: Mean per year does not include 'Combined years' or ANZ

Pearson correlation analyses – General Surgery – Intra-assessment: Tables

Selection items

Table 44 *Correlations – GS ANZ selection performance for 2008, 2009 and 2010 selection years*

		2008						**2009**						**2010**						**Combined years**					
		CV%			**RR%**			**CV%**			**RR%**			**CV%**			**RR%**			**CV%**			**RR%**		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
RR%	*r*	.13	.02	.11	–	–	–	-.16	.10	-.13	–	–	–	-.14	-.17	-.13	–	–	–	-.02	.06	-.01	–	–	–
	n	81	19	100				89	18	107				121	19	140				291	56	347			
Int%	*r*	-.19	.14	-.15	.06	-.37	.03	**-.38****	.00	**-.34****	-.08	.06	-.06	**-.22***	.32	-.17	**-.18***	.09	**-.21***	**-.26****	.02	**-.23****	-.02	**-.38****	-.07
	n	81	19	100	81	19	100	89	18	107	89	18	107	121	19	140	121	19	140	291	56	347	291	56	347

CV = Curriculum Vitae; RR = Referee Report; Int = Interview

r = Pearson Correlation; *n* = number of candidates

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

1.

Examinations

Table 45 *Correlations – GS ANZ examination performance for 2008, 2009 and 2010 selection years*

		2008						2009						2010						Combined years					
		GenSSE%			SpecSSE%			GenSSE%			SpecSSE%			GenSSE%			SpecSSE%			GenSSE%			SpecSSE%		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
SpecSSE%	*r*	**.82****	**.86****	**.82****	–	–	–	**.86****	**.89****	**.86****	–	–	–	**.86****	**.92****	**.87****	–	–	–	**.85****	**.89****	**.86****	–	–	–
	n	56	10	66				80	13	93				114	18	132				250	41	291			
CE%	*r*	**.30****	.46	**.33****	**.42****	.18	**.39****	**.35****	**.79****	**.42****	**.30****	**.81****	**.34****	**.45****	.21	**.40****	**.44****	.38	**.41****	**.34****	**.49****	**.36****	**.35****	**.40***	**.35****
	n	78	18	96	56	10	66	87	17	104	80	13	93	114	19	133	110	18	128	279	54	333	246	41	287

Gen% = Generic SSE score; Spec% = Specialty SSE score; CE% = Clinical Exam score

r = Pearson Correlation; *n* = number of candidates

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

2.

Direct Observation of Procedural Skills (DOPS)

Table 46 *Correlations – GS ANZ DOPS performance for 2008 and 2009 selection years, n ≥ 5*

		DOPS1%			DOPS2%			DOPS3%			DOPS4%			Average DOPS%		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
										2009						
DOPS1%	*r*	_	_	_	.32*	-.02	.28*	.02	Y	.08	.24	Y	.150	.76**	.68*	.74**
	n				59	8	67	30	2	32	13	2	15	64	10	74
DOPS2%	*r*	.46**	.21	.43**	_	_	_	.23	Y	.28	.46	Y	.37	.80**	.86**	.79**
	n	39	11	50				30	2	32	13	2	15	61	8	69
DOPS3%	*r*	-.06	Y	-.01	.30	Y	.34	_	_	_	.20	Y	.03	.53**	Y	.56**
	n	20	2	22	20	2	22				13	2	15	30	2	32
DOPS4%	*r*	-.20	a	-.20	.09	a	.09	.36	a	.36	_	_	_	.70**	Y	.58**
	n	10	0	10	10	0	10	10	0	10				13	2	15
Average DOPS%	*r*	.75**	.82**	.78**	.78**	.78**	.79**	.63**	Y	.66**	.56	a	.56	_	_	_
	n	45	14	59	39	11	50	20	2	22	10	0	10			
								2008								

Note: 2008 is below the diagonal, 2009 is above the diagonal

r = Pearson Correlation, *n* = number of candidates

a. Cannot be computed because at least one of the variables is constant.

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

Y indicates no scores listed because n < 5 (see Assessments including n < 5
Table 52 Correlations – GS ANZ DOPS performance for 2008 and 2009 selection years, including n < 5)

Table 47 *Correlations – GS ANZ DOPS performance for 2010 and combined selection years, n ≥ 5*

		DOPS1%			DOPS2%			DOPS3%			DOPS4%			Average DOPS%		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
					2008, 2009, 2010 combined											
DOPS1%	*r*	_	_	_	**.28****	.04	**.25****	.04	Y	.07	.10	Y	.07	**.76****	**.79****	**.76****
	n	102	17	119	174	27	201	84	4	88	40	2	42	211	41	252
DOPS2%	*r*	.15	-.42	.10	_	_	_	**.27***	Y	**.30****	.31	Y	.30	**.78****	**.74****	**.77****
	n	76	8	84				84	4	88	40	2	42	177	27	204
DOPS3%	*r*	.10	a	.10	.29	a	.29	_	_	_	.21	Y	.16	**.61****	Y	**.62****
	n	34	0	34	34	0	34				39	2	41	84	4	88
DOPS4%	*r*	.12	a	.12	.33	a	.33	.12	a	.12	_	_	_	**.62****	Y	**.57****
	n	17	0	17	17	0	17	16	0	16				40	2	42
Average DOPS%	*r*	**.76****	**.80****	**.76****	**.76****	.58	**.75****	**.64****	a	**.64****	**.59***	a	**.59****	_	_	_
	n	102	17	119	77	8	85	34	0	34	17	0	17			
							2010									

Note: 2010 is below the diagonal, combined years is above the diagonal

r = Pearson Correlation, *n* = number of candidates

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

Y indicates no scores listed because n < 5 (see

Table *53 Correlations – GS ANZ DOPS performance for 2010 and combined selection years, including n* < 5

Mini Clinical Evaluation Exercise (MiniCEX)

Table 48 *Correlations – GS ANZ MiniCEX performance for 2008 and 2009 selection years, n ≥ 5*

		MiniCEX1%			MiniCEX2%			MiniCEX3%			MiniCEX4%			Average MiniCEX%		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
					2009											
MiniCEX1%	*r*	_	_	_	.13	.06	.13	**.55****	Y	**.52****	.41	Y	.38	**.84***	**.74***	**.83****
	n				58	8	66	23	2	25	10	2	12	67	10	77
MiniCEX2%	*r*	.08	.12	.10	_	_	_	.32	Y	.37	**.76***	Y	**.74****	**.69***	**.72***	**.70****
	n	40	10	50				23	2	25	10	2	12	58	8	66
MiniCEX3%	*r*	.34	Y	.35	.04	Y	.18	_	_	_	.43	Y	.43	**.78****	Y	**.79****
	n	20	2	22	20	2	22				10	2	12	24	2	26
MiniCEX4%	*r*	-.21	a	-.21	-.34	a	-.34	.64	a	.64	_	_	_	**.89****	Y	**.86****
	n	9	0	9	9	0	9	9	0	9				10	2	12
Average MiniCEX%	*r*	**.66****	**.81****	**.69****	**.63****	**.80****	**.65****	**.74****	Y	**.77****	.61	a	.61	_	_	_
	n	46	14	60	41	10	51	20	2	22	9	0	9			
							2008									

Note: 2008 is below the diagonal, 2009 is above the diagonal

r = Pearson Correlation, *n* = number of candidates

a. Cannot be computed because at least one of the variables is constant.

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

Y indicates no scores listed because n < 5 (see Table 54 Correlations – GS ANZ MiniCEX performance for 2008 and 2009 selection years, including n < 5)

Table 49 *Correlations – GS ANZ MiniCEX performance for 2010 and combined selection years, n ≥ 5*

		MiniCEX1%			MiniCEX2%			MiniCEX3%			MiniCEX4%			Average MiniCEX%		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
					2008, 2009, 2010 combined											
MiniCEX1%	*r*	_	_	_	.13	.13	.12	**.30****	Y	**.31****	.03	Y	.03	**.78****	**.80****	**.78****
	n				171	26	197	76	4	80	30	2	32	213	41	254
MiniCEX2%	*r*	.15	.23	.13	_	_	_	.19	Y	**.23***	.13	Y	.15	**.71****	**.77****	**.72****
	n	73	8	81				77	2	81	31	2	33	173	26	199
MiniCEX3%	*r*	.08	a	.08	.16	a	.16	_	_	_	**.41***	Y	**.42***	**.68****	Y	**.70****
	n	33	0	33	34	0	34				31	2	33	78	4	82
MiniCEX4%	*r*	.04	a	.04	.40	a	.40	.34	a	.34	_	_	_	**.65****	Y	**.65****
	n	11	0	11	12	0	12	12	0	12				31	2	33
Average MiniCEX%	*r*	**.78****	**.83****	**.78****	**.76****	**.78***	**.76****	**.59****	a	**.59****	**.74****	a	**.74****	_	_	_
	n	100	17	117	74	8	82	34	0	34	12	0	12			
		2010														

Note: 2010 is below the diagonal, combined years is above the diagonal

r = Pearson Correlation, *n* = number of candidates

a. Cannot be computed because at least one of the variables is constant.

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

Y indicates no scores listed because n < 5 (see Table 55 Correlations – GS ANZ MiniCEX performance for 2010 and combined selection years, including n < 5)

End of Term Assessments (ETA)

Table 50 *Correlations – GS ANZ ETA performance for 2008 and 2009 selection years, n ≥ 5*

		ETA1%			**ETA2%**			**ETA3%**			**ETA4%**			**Average ETA%**		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
					2009											
ETA1%	*r*	_	_	_	**.29***	**.62***	**.34****	.24	.55	**.25***	.16	.41	.20	**.70****	**.87****	**.73****
	n				73	13	86	66	7	73	54	13	67	77	16	93
ETA2%	*r*	-.06	.08	-.03	_	_	_	.05	.02	.06	**.42****	**.67***	**.45****	**.59****	**.86****	**.63****
	n	58	19	77				65	6	71	53	13	66	75	15	90
ETA3%	*r*	.15	.21	.16	.16	.44	.20	_	_	_	.22	.47	.24	**.67****	.74	**.68****
	n	53	19	72	53	19	72				51	7	58	68	7	75
ETA4%	*r*	.12	.30	.15	.12	**.48***	.18	.25	.45	**.30***	_	_	_	**.72****	**.77****	**.71****
	n	48	19	67	49	19	68	49	19	68				56	14	70
Average ETA%	*r*	**.38****	**.58****	**.42****	**.48****	**.70****	**.52****	**.55****	**.75****	**.58****	**.62****	**.80****	**.66****	_	_	_
	n	59	19	78	60	19	79	55	19	74	50	19	69			
								2008								

Note: 2008 is below the diagonal, 2009 is above the diagonal

r = Pearson Correlation, *n* = number of candidates

a. Cannot be computed because at least one of the variables is constant.

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

Table 51 *Correlations - GS ANZ ETA performance for 2010 and combined selection years, n ≥ 5*

		ETA1%			ETA2%			ETA3%			ETA4%			Average ETA%		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
					2008, 2009, 2010 combined											
ETA1%	*r*	_	_	_	.13	**.32***	**.16****	**.20***	.28	**.21***	.14	.32	**.17***	**.66****	**.75****	**.68****
	n	90		106	215	48	263	119	26	145	102	32	134	226	51	277
ETA2%	*r*	.14	.33	.18	_	_	_	.09	**.47***	.12	**.27****	**.57****	**.31****	**.65****	**.81****	**.68****
	n	84	16	100				118	25	143	102	32	134	245	52	297
ETA3%	*r*	a	a	a	a	a	a	_	_	_	**.24***	**.47***	**.27****	**.62****	**.76****	**.63****
	n	0	0	0	0	0	0				100	26	126	123	26	149
ETA4%	*r*	a	a	a	a	a	a	a	a	a	_	_	_	**.68****	**.78****	**.69****
	n	0	0	0	0	0	0	0	0	0				106	33	139
Average ETA%	*r*	**.79****	**.80****	**.79****	**.81****	**.86****	**.82****	a	a	a	a	a	a	_	_	_
	n	90	16	106	110	18	128	0	0	0	0	0	0			
								2010								

Note: 2010 is below the diagonal, combined years is above the diagonal

r = Pearson Correlation, *n* = number of candidates

a. Cannot be computed because at least one of the variables is constant.

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

Assessments including n < 5

Table 52 *Correlations – GS ANZ DOPS performance for 2008 and 2009 selection years, including n < 5*

		DOPS1%			DOPS2%			DOPS3%			DOPS4%			Average DOPS%		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
					2009											
DOPS1%	*r*	_	_	_	**.32***	-.02	**.28***	.02	**1.00****	.08	.24	**-1.00****	.15	**.76****	**.68***	**.74****
	n				59	8	67	30	2	32	13	2	15	64	10	74
DOPS2%	*r*	**.46****	.21	**.43****	_	_	_	.23	**1.00****	.28	.46	**-1.00****	.37	**.80****	**.86****	**.79****
	n	39	11	50				30	2	32	13	2	15	61	8	69
DOPS3%	*r*	-.06	**1.00****	-.01	.30	**1.00****	.34	_	_	_	.20	**-1.00****	.03	**.54****	**1.00****	**.55****
	n	20	2	22	20	2	22				13	2	15	30	2	32
DOPS4%	*r*	-.20	a	-.20	.09	a	.09	.36	a	.36	_	_	_	**.70****	**-1.00****	**.58****
	n	10	0	10	10	0	10	10	0	10				13	2	15
Average DOPS%	*r*	**.76****	**.82****	**.78****	**.78****	**.78****	**.79****	**.63****	**1.00****	**.66****	.56	a	.56	_	_	_
	n	45	14	59	39	11	50	20	2	22	10	0	10			
							2008									

Note: 2008 is below the diagonal, 2009 is above the diagonal

r = Pearson Correlation, *n* = number of candidates

a. Cannot be computed because at least one of the variables is constant.

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

Table 53 *Correlations – GS ANZ DOPS performance for 2010 and combined selection years, including n < 5*

		DOPS1%			**DOPS2%**			**DOPS3%**			**DOPS4%**			**Average DOPS%**		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
					2008, 2009, 2010 combined											
DOPS1%	*r*	–	–	–	**.28****	.04	**.25****	.04	.66	.07	.10	**-1.00****	.07	**.76****	**.79****	**.76****
	n				174	27	201	84	4	88	40	2	42	211	41	252
DOPS2%	*r*	.15	-.42	.12	–	–	–	**.27***	**.97***	**.30****	.31	**-1.00****	.26	**.78****	**.74****	**.77****
	n	76	8	84				84	4	88	40	2	42	177	27	204
DOPS3%	*r*	.10	a	.10	.29	C	.29	–	–	–	.21	**-1.00****	.16	**.61****	.91	**.62****
	n	34	0	34	34	0	34				39	2	41	84	4	88
DOPS4%	*r*	.12	a	.12	.33	a	.33	.12	a	.12	–	–	–	**.62****	**-1.00****	**.57****
	n	17	0	17	17	0	17	16	0	16				40	2	42
Average DOPS%	*r*	**.76****	**.80****	**.76****	**.76****	.58	**.75****	**.64****	a	**.64****	**.59***	a	**.59****			
	n	102	17	119	77	8	85	34	0	34	17	0	17			
		2010														

Note: 2010 is below the diagonal, combined years above the diagonal

r = Pearson Correlation, *n* = number of candidates

a. Cannot be computed because at least one of the variables is constant.

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

Table 54 *Correlations – GS ANZ MiniCEX performance for 2008 and 2009 selection years, including n < 5*

		MiniCEX1%			MiniCEX2%			MiniCEX3%			MiniCEX4			Average MiniCEX%		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
					2009											
MiniCEX1%	*r*	_	_	_	.13	.06	.13	**.55****	**1.00****	**.52****	.41	**1.00****	.38	**.84***	**.74***	**.83****
	n				58	8	66	23	2	25	10	2	12	67	10	77
MiniCEX2%	*r*	.079	.116	.079	_	_	_	.32	**1.00****	.37	**.76***	**1.00****	**.74****	**.69***	**.72***	**.70****
	n	40	10	50				23	2	25	10	2	12	58	8	66
MiniCEX3%	*r*	.34	**1.00****	.35	.04	**1.00****	.18	_	_	_	.43	**1.00****	.43	**.78****	**1.00****	**.79****
	n	20	2	22	20	2	22				10	2	12	24	2	26
MiniCEX4%	*r*	-.21	a	-.21	-.34	a	-.34	.64	a	.64	_	_	_	**.89****	**1.00****	**.86****
	n	9	0	9	9	0	9	9	0	9				10	2	12
Average	*r*	**.66****	**.81****	**.69****	**.63****	**.80****	**.65****	**.74****	**1.00****	**.77****	.61	a	.61	_	_	_
MiniCEX%	*n*	46	14	60	41	10	51	20	2	22	9	0	9			
								2008								

Note: 2008 is below the diagonal, 2009 is above the diagonal

r = Pearson Correlation, *n* = number of candidates

a. Cannot be computed because at least one of the variables is constant.

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

Table 55 *Correlations – GS ANZ MiniCEX performance for 2010 and combined selection years, including n < 5*

		MiniCEX1%			**MiniCEX2%**			**MiniCEX3%**			**MiniCEX4%**			**Average MiniCEX%**		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
					2008, 2009, 2010 combined											
MiniCEX1%	*r*	–	–	–	.13	.13	.12	**.30****	.42	**.31****	.03	**1.00****	.03	**.78****	**.80****	**.78****
	n				171	26	197	76	4	80	30	2	32	213	41	254
MiniCEX2%	*r*	.15	.23	.13	–	–	–	.19	**1.00****	**.23***	.13	**1.00****	.15	**.71****	**.77****	**.72****
	n	73	8	81				77	2	81	31	2	33	173	26	199
MiniCEX3%	*r*	.08	a	.08	.16	a	.16	–	–	–	**.41***	**1.00****	**.42***	**.68****	.95	**.70****
	n	33	0	33	34	0	34				31	2	33	78	4	82
MiniCEX4%	*r*	.04	a	.04	.40	a	.40	.34	a	.34	–	–	–	**.65****	**1.00****	**.65****
	n	11	0	11	12	0	12	12	0	12				31	2	33
Average	*r*	**.78****	**.83**	**.78**	**.76****	**.78***	**.76****	**.59****	a	**.59****	**.74****	a	**.74****	–	–	–
MiniCEX%	*n*	100	17	117	74	8	82	34	0	34	12	0	12			
						2010										

Note: 2010 is below the diagonal, combined years above the diagonal

r = Pearson Correlation, *n* = number of candidates

a. Cannot be computed because at least one of the variables is constant.

** Correlation is significant at .01 (2-tailed)

* Correlation is significant at .05 (2-tailed)

Correlations possible and identified

Table 56 *Number of correlations possible from available data*

	2008			2009			2010			Combined years			
Country	**Au**	**NZ**	**ANZ**	**Au**	**NZ**	**ANZ**	**Au**	**NZ**	**ANZ**	**Au**	**NZ**	**ANZ**	
Assessment	**Number of correlations possible**												
GenSSE	4	4	4	4	4	4	4	4	4	4	4	4	
SpecSSE	4	4	4	4	4	4	4	4	4	4	4	4	
CE	4	4	4	4	4	4	4	4	4	4	4	4	**144**
DOPS1	4	4	4	4	4	4	4	4	4	4	4	4	
DOPS2	4	4	4	4	4	4	4	4	4	4	4	4	
DOPS3	4	-	4	4	-	4	4	-	4	4	-	4	
DOPS4	4	-	4	4	-	4	4	-	4	4	-	4	
Av DOPS	4	4	4	4	4	4	4	4	4	4	4	4	**208**
MiniCEX1	4	4	4	4	4	4	4	4	4	4	4	4	
MiniCEX2	4	4	4	4	4	4	4	4	4	4	4	4	
MiniCESX3	4	-	4	4	-	4	4	-	4	4	-	4	
MiniCEX4	4	-	4	4	-	4	4	-	4	4	-	4	
Av MiniCEX	4	4	4	4	4	4	4	4	4	4	4	4	**208**
ETA1	4	4	4	4	4	4	4	4	4	4	4	4	
ETA2	4	4	4	4	4	4	4	4	4	4	4	4	
ETA3	4	4	4	4	4	4	-	-	-	4	4	4	
ETA4	4	4	4	4	4	4	-	-	-	4	4	4	
Av ETA	4	4	4	4	4	4	4	4	4	4	4	4	**216**
	72	**56**	**72**	**72**	**56**	**72**	**64**	**48**	**64**	**72**	**56**	**72**	**776**

Note: - = no data available, or n ≤ 5

Table 57 *Number of significant correlations identified*

	2008			2009			2010			Combined years			
Country	**Au**	**NZ**	**ANZ**	**Au**	**NZ**	**ANZ**	**Au**	**NZ**	**ANZ**	**Au**	**NZ**	**ANZ**	
Assessment	**Number of correlations identified**												
GenSSE	0	0	0	1	1	1	0	2	0	0	3	0	
SpecSSE	0	0	0	2	1	2	0	2	1	1	3	1	
CE	0	0	0	4	1	3	2	1	2	3	0	3	**40**
DOPS1	1	2	2	0	1	0	3	0	3	2	0	1	
DOPS2	0	2	1	1	1	2	0	1	0	2	1	1	
DOPS3	0	-	1	0	-	0	1	-	1	0	-	0	
DOPS4	0	-	0	1	-	0	0	-	0	1	-	0	
Av DOPS	1	1	1	2	1	2	1	0	1	2	1	1	**46**
MiniCEX1	1	0	1	1	0	1	1	0	2	0	1	1	
MiniCEX2	0	1	0	1	0	1	0	0	0	0	1	0	
MiniCESX3	0	-	0	0	-	0	0	-	0	0	-	0	
MiniCEX4	0	-	0	0	-	1	0	-	0	0	-	0	
Av MiniCEX	0	1	0	2	0	2	1	0	1	1	2	2	**26**
ETA1	0	0	0	2	1	3	2	0	2	2	1	2	
ETA2	1	0	1	2	2	3	0	1	0	0	0	0	
ETA3	1	1	1	1	1	3	-	-	-	2	2	2	
ETA4	2	1	3	1	2	2	-	-	-	3	2	2	
Av ETA	0	1	0	2	3	3	1	0	1	2	2	2	**74**
	7	**10**	**11**	**23**	**15**	**29**	**12**	**7**	**14**	**21**	**19**	**18**	**186**

Note: - = no data available, or n ≤ 5

Pearson correlation analyses – very weak to very strong

Table 58 *Summary of very weak significant Pearson correlations (r .1 to .15) by year and selection item*

Year	Country	Selection item	Assessment item	Strength of correlation (*r*)	Number of trainees
2010	ANZ	RR	DOPS1	.15	119
2010	ANZ	Total sel	SpecSSE	.14	132
2010	ANZ	Total sel	MiniCEX1	.15	117
Combined years	ANZ	CV	Av MiniCEX	-.12	257
Combined years	Au	CV	Av ETA	-.12	257
Combined years	ANZ	CV	Av ETA	-.10	312
Combined years	Au	RR	DOPS1	.14	211
Combined years	ANZ	RR	Av DOPS	.13	255
Combined years	Au	RR	Av MiniCEX	.13	216
Combined years	ANZ	RR	Av MiniCEX	.13	257
Combined years	ANZ	RR	ETA1	.13	277
Combined years	Au	Int	SpecSSE	.10	250
Combined years	ANZ	Int	SpecSSE	.11	291
Combined years	Au	Int	DOPS2	.12	177
Combined years	ANZ	Int	MiniCEX1	.12	254
Combined years	ANZ	Total sel	DOPS1	.13	252
Combined years	Au	Total sel	Av DOPS	.14	214
Combined years	Au	Total sel	ETA1	.13	226
Combined years	ANZ	Total sel	ETA1	.11	277

Table 59 *Summary of weak Pearson Correlations (r .16 to .29) by year and selection item*

Year	Country	Selection item	Assessment item	Strength of correlation (*r*)	Number of trainees
2008	Au	CV	Av DOPS	-.27	45
2008	Au	CV	DOPS1	-.25	45
2008	ANZ	CV	DOPS2	-.28	50
2008	Au	CV	ETA3	-.28	55
2008	ANZ	CV	ETA3	-.25	74
2008	ANZ	Int	DOPS1	.22	59
2008	Au	Int	ETA2	-.27	60
2008	ANZ	Int	ETA2	-.21	79
2008	ANZ	Int	ETA4	.20	69
2008	ANZ	RR	ETA4	.24	69
2008	ANZ	Total sel	ETA4	.24	69
2008	Au	Total sel	ETA4	.24	50
2009	Au	CV	SpecSSE	-.23	80
2009	ANZ	CV	SpecSSE	-.21	93
2009	Au	CV	CE	-.22	88
2009	ANZ	CV	CE	-.18	105
2009	Au	CV	MiniCEX1	-.22	67
2009	ANZ	CV	MiniCEX1	-.19	77
2009	Au	CV	Av MiniCEX	-.27	68
2009	ANZ	CV	Av MiniCEX	-.24	78
2009	ANZ	CV	ETA3	-.21	75
2009	Au	RR	CE	.19	88
2009	Au	RR	DOPS2	.27	61
2009	ANZ	RR	DOPS2	.25	69
2009	Au	RR	Av DOPS	.26	66
2009	ANZ	RR	Av DOPS	.20	76
2009	Au	RR	MiniCEX2	.25	58
2009	ANZ	RR	MiniCEX2	.23	66
2009	Au	RR	Av MiniCEX	.26	68
2009	ANZ	RR	Av MiniCEX	.25	78
2009	Au	RR	ETA1	.29	77
2009	ANZ	RR	ETA1	.27	93
2009	Au	RR	ETA2	.20	75
2009	ANZ	RR	ETA2	.24	90
2009	Au	Int	GenSSE	.19	88
2009	ANZ	Int	GenSSE	.20	105

Year	Country	Selection item	Assessment item	Strength of correlation (*r*)	Number of trainees
2009	Au	Int	SpecSSE	.21	80
2009	ANZ	Int	SpecSSE	.26	93
2009	ANZ	Int	ETA1	.19	93
2009	ANZ	Int	ETA2	.21	90
2009	ANZ	Int	Av ETA	.19	97
2009	Au	Total sel	CE	.28	88
2009	ANZ	Total sel	CE	.27	105
2009	ANZ	Total sel	DOPS2	.22	69
2009	ANZ	Total sel	Av DOPS	.23	76
2009	Au	Total sel	Av DOPS	.27	66
2009	Au	Total sel	ETA1	.29	77
2009	Au	Total sel	ETA2	.26	75
2009	ANZ	Total sel	ETA3	.19	75
2009	ANZ	Total sel	ETA4	.23	70
2010	ANZ	CV	DOPS1	.22	119
2010	Au	CV	DOPS1	.23	102
2010	Au	RR	DOPS1	.17	102
2010	Au	RR	Av DOPS	.18	103
2010	ANZ	RR	Av DOPS	.18	120
2010	Au	RR	MiniCEX1	.21	100
2010	ANZ	RR	MiniCEX1	.21	117
2010	Au	RR	Av MiniCEX	.18	101
2010	ANZ	RR	Av MiniCEX	.19	118
2010	Au	RR	ETA1	.28	90
2010	ANZ	RR	ETA1	.25	106
2010	Au	RR	Av ETA	.19	116
2010	ANZ	RR	Av ETA	.22	134
2010	Au	Int	CE	.18	116
2010	ANZ	Int	CE	.19	135
2010	Au	Total sel	CE	.22	116
2010	ANZ	Total sel	CE	.19	135
2010	Au	Total sel	DOPS1	.29	102
2010	ANZ	Total sel	DOPS1	.25	119
2010	ANZ	Total sel	ETA1	.23	106
Combined years	NZ	CV	GenSSE	.25	54
Combined years	Au	CV	DOPS2	-.16	177
Combined years	ANZ	CV	DOPS2	-.19	204
Combined years	Au	CV	ETA3	-.23	123

Year	Country	Selection item	Assessment item	Strength of correlation (r)	Number of trainees
Combined years	ANZ	CV	ETA3	-.23	149
Combined years	Au	CV	ETA4	-.17	106
Combined years	NZ	RR	SpecSSE	.28	41
Combined years	Au	RR	CE	.22	283
Combined years	ANZ	RR	CE	.21	337
Combined years	Au	RR	DOPS4	.27	40
Combined years	Au	RR	Av DOPS	.19	214
Combined years	Au	RR	ETA1	.16	226
Combined years	Au	RR	ETA3	.27	123
Combined years	ANZ	RR	ETA3	.26	149
Combined years	ANZ	RR	Av ETA	.19	312
Combined years	Au	RR	Av ETA	.19	257
Combined years	NZ	Int	SpecSSE	.29	41
Combined years	ANZ	Int	CE	.20	337
Combined years	Au	Int	CE	.20	283
Combined years	NZ	Int	ETA1	.29	51
Combined years	NZ	Int	ETA4	.28	33
Combined years	NZ	Int	Av ETA	.24	55
Combined years	ANZ	Total sel	CE	.24	337
Combined years	Au	Total sel	CE	.24	283
Combined years	Au	Total sel	DOPS1	.19	211
Combined years	NZ	Total sel	ETA3	.28	26
Combined years	Au	Total sel	ETA4	.20	106
Combined years	NZ	Total sel	ETA4	.29	33
Combined years	ANZ	Total sel	ETA4	.23	139
Combined years	NZ	Total sel	Av ETA	.25	55

Table 60 *Summary of moderate Pearson Correlations (r .30 to .49) by year and selection item*

Year	Country	Selection item	Assessment item	Strength of correlation (*r*)	Number of trainees
2008	ANZ	CV	DOPS3	-.36	22
2008	ANZ	CV	Av DOPS	-.32	59
2008	ANZ	CV	DOPS1	-.30	59
2008	ANZ	Int	MiniCEX1	.31	60
2008	Au	Int	MiniCEX1	.30	46
2008	Au	RR	ETA4	.30	50
2008	NZ	Int	ETA4	.38	19
2008	NZ	Int	ETA3	.43	19
2008	NZ	Int	Av ETA	.49	19
2009	Au	RR	ETA3	.41	68
2009	ANZ	RR	ETA3	.37	75
2009	Au	RR	ETA4	.40	56
2009	ANZ	RR	ETA4	.42	70
2009	Au	RR	Av ETA	.48	79
2009	NZ	RR	Av ETA	.39	18
2009	ANZ	RR	Av ETA	.46	97
2009	NZ	Int	GenSSE	.43	17
2009	Au	Int	CE	.39	88
2009	NZ	Int	CE	.46	17
2009	ANZ	Int	CE	.38	105
2009	ANZ	Total sel	ETA1	.30	93
2009	ANZ	Total sel	ETA2	.33	90
2009	Au	Total sel	Av ETA	.32	79
2009	ANZ	Total sel	Av ETA	.36	97
2010	NZ	CV	CE	-.48	19
2010	Au	Int	DOPS3	-.48	34
2010	ANZ	Int	DOPS3	-.48	34
2010	Au	Total sel	ETA1	.30	90
2010	NZ	RR	ETA2	.43	18
Combined years	NZ	CV	DOPS2	-.48	27
Combined years	NZ	CV	Av DOPS	-.35	41
Combined years	NZ	CV	MiniCEX2	-.39	26
Combined years	NZ	CV	Av MiniCEX	-.33	41
Combined years	ANZ	RR	ETA4	.30	139
Combined years	Au	RR	ETA4	.33	106

Year	Country	Selection item	Assessment item	Strength of correlation (*r*)	Number of trainees
Combined years	NZ	Int	GenSSE	.32	54
Combined years	NZ	Int	MiniCEX1	.31	41
Combined years	NZ	Int	Av MiniCEX	.31	41
Combined years	NZ	Total sel	GenSSE	.44	54
Combined years	NZ	Total sel	SpecSSE	.48	41

Table 61 *Summary of strong Pearson correlations (r .50 to .69) by year and selection item*

Year	Country	Selection item	Assessment item	Strength of correlation (*r*)	Number of trainees
2008	NZ	CV	DOPS1	-.62	14
2008	NZ	CV	DOPS2	-.63	11
2008	NZ	CV	MiniCEX2	-.58	10
2008	NZ	CV	Av MiniCEX	-.60	14
2008	NZ	Int	DOPS1	.50	14
2008	NZ	Total sel	DOPS2	-.52	11
2009	ANZ	CV	MiniCEX4	-.50	12
2009	NZ	RR	Average DOPS	-.60	10
2009	NZ	RR	ETA2	.55	15
2009	NZ	RR	ETA4	.66	14
2009	NZ	Int	SpecSSE	.67	13
2009	NZ	Int	DOPS2	.64	8
2009	Au	Int	DOPS4	-.59	13
2009	NZ	Int	ETA1	.60	16
2009	NZ	Int	Average ETA	.57	18
2009	NZ	Total sel	ETA2	.65	15
2009	NZ	Total sel	ETA4	.57	14
2009	NZ	Total sel	Average ETA	.55	18
2010	NZ	Int	GenSSE	.68	19
2010	NZ	Int	SpecSSE	.69	18
2010	NZ	Total sel	GenSSE	.69	19
2010	NZ	Total sel	SpecSSE	.54	18
Combined years	NZ	Int	ETA3	.54	26

Table 62 *Summary of very strong significant Pearson correlations (r ≥ .7) by year and selection item*

Year	**Country**	**Selection item**	**Assessment item**	**Strength of correlation (*r*)**	**Number of trainees**
2008	NZ	CV	Average DOPS	-.70	14
2009	NZ	RR	DOPS1	-.73	10
2009	NZ	Int	ETA3	-.80	7
2010	NZ	CV	DOPS2	-.75	8

Table 63 *Summary of very strong significant Pearson Correlations (r ≥ .7) for n ≤ 5 by year and selection item*

Year	**Country**	**Selection item**	**Assessment item**	**Strength of correlation (*r*)**	**Number of trainees**
Combined years	NZ	CV	DOPS3	-.97	4
Combined years	NZ	Total sel	DOPS3	-.95	4
Combined years	NZ	CV	MiniCEX3	-.99	4
Combined years	NZ	Total sel	MiniCEX3	-.97	4

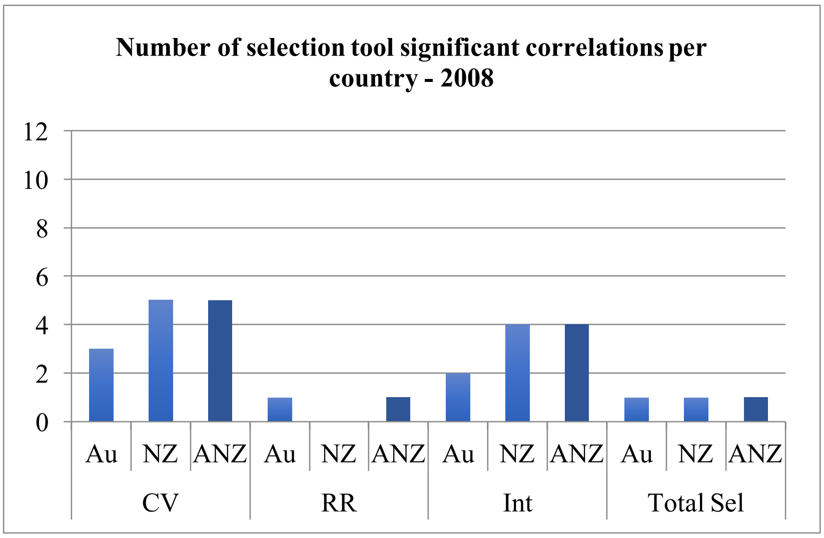

Figure 37. Selection tool significant Pearson correlations per country - 2008

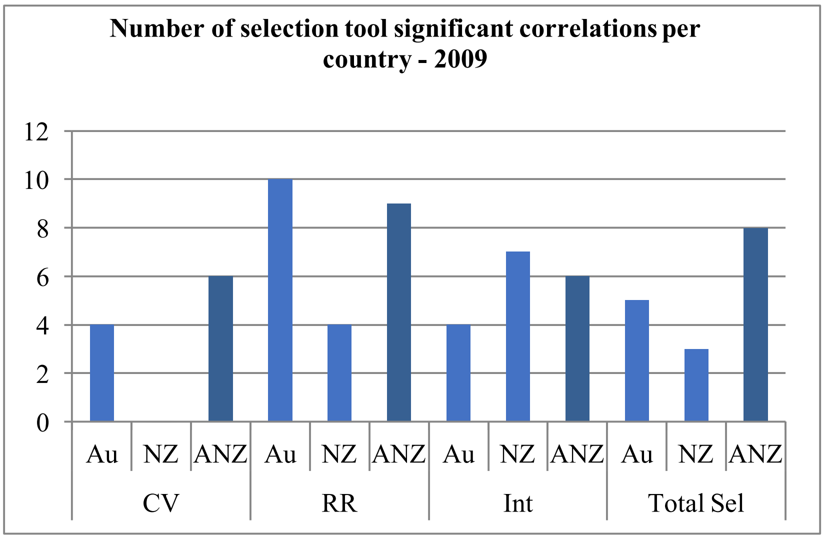

Figure 38. Selection tool significant Pearson correlations per country - 2009

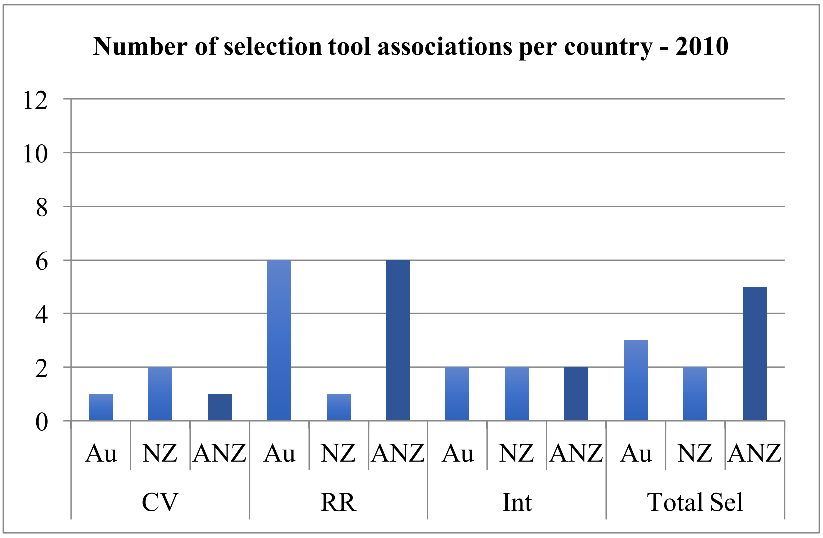

Figure 39. Selection tool significant Pearson correlations per country - 2010

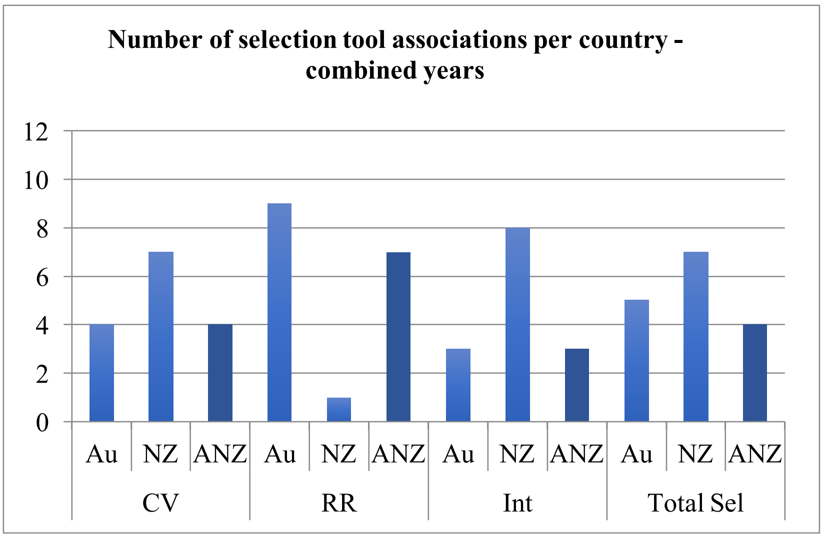

Figure 40. Selection tool significant Pearson correlations per country - combined years

Table 64 *Significant correlations between GS ANZ selection performance and DOPS performance for 2008, 2009, 2010 and combined selection years; including n < 5*

		Pearson Correlations (*r*)														
		DOPS1 %			DOPS2 %			DOPS3 %			DOPS4 %			Average DOPS %		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
2008																
	CV %	-.25*	-.62**	-.30**	-.23	-.63*	-.28*	-.30	-1.00**	-.36*	.22	_	.22	-.27*	-.70**	-.32**
	RR %	.03	-.10	-.00	.10	.20	.09	.04	1.00**	.04	.17	_	.17	.07	.05	.05
	Int %	.22	.50*	.22*	.15	-.33	.06	-.01	1.00**	-.00	-.07	_	-.07	.06	.26	.05
	Total sel %	.08	-.31	.02	.07	-.52*	-.04	-.12	-1.00**	-.15	.19	_	.19	-.02	-.39	-.08
N	100	45	14	59	39	11	50	20	2	22	10	0	10	45	14	59
2009																
	CV %	-.02	-.11	-.03	-.12	-.19	-.12	.20	-1.00**	.12	-.07	1.00**	.00	.03	-.09	.02
	RR %	.18	-.73**	.09	.27*	-.10	.25*	.22	-1.00**	.22	.43	1.00**	.31	.26*	-.60*	.20*
	Int %	.08	.24	.10	.13	.64*	.17	-.12	1.00**	-.11	-.59*	-1.00**	-.40	.08	.46	.12
	Total sel %	.17	-.18	.11	.20	.39	.22*	.28	-1.00**	.22	-.09	-1.00**	.-.03	.27*	.01	.23*
N	107	64	10	74	61	8	69	30	2	32	13	2	15	66	10	76
2010																
	CV %	.23**	.09	.22**	-.11	-.75*	-.13	-.01		-.01	-.33	_	-.33	.04	-.16	.03
	RR %	.17*	.11	.15*	.02	.36	.03	.23	_	.23	.38	_	.38	.18*	.32	.18*
	Int %	.00	-.26	.01	.09	-.30	.06	-.48**	_	-.48**	-.02	_	-.02	.00	-.31	.00
	Total sel %	.29**	-.08	.25**	-.03	-.26	-.05	-.07	_	-.07	-.04	_	-.04	.15	-.11	.13
N	140	102	17	119	77	8	85	34	0	34	17	0	17	103	17	120
Combined years																
	CV %	.06	-.23	.03	-.16*	-.48**	-.19**	-.02	-.97**	-.06	-.14	1.00**	-.09	-.01	-.35**	-.05
	RR %	.14*	-.30	.08	.12	.02	.10	.14	-.06	.13	.27*	1.00**	.24	.19**	-.21	.13*
	Int %	.09	.18	.09	.12*	.15	.10	-.17	.39	-.16	-.18	-1.00**	-.15	.04	.21	.05
	Total sel %	.19**	-.24	.13*	.06	-.20	.02	-.03	-.95*	-.06	.00	-1.00**	.04	.14*	-.21	.09
N	347	211	41	252	177	27	204	84	4	88	40	2	42	214	41	255

Notes. CV% = CV score as a percentage. RR% = Referee reports score as a percentage. Int % = Interview score as a percentage. Total sel % = Total selection score as a percentage. DOPS % = DOPS Assessment score. Average DOPS % = Mean score of DOPS 1-4.
r = Pearson Correlation. N = Number of trainees.
** = Correlation is significant at .01 (1-tailed). * = Correlation is significant at .05 (1-tailed). Teal = Also significant in regression analysis.

Table 65 *Significant correlations between GS ANZ selection and MiniCEX performance for 2008, 2009, 2010 selection years and combined selection years; including n < 5*

		Pearson Correlations (*r*)														
		MiniCEX1 %			MiniCEX 2 %			MiniCEX 3 %			MiniCEX 4 %			Average MiniCEX %		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
2008																
	CV %	-.11	-.38	-.18	-.03	-.58*	-.15	-.17	-1.00**	-.29	-.04	_	-.04	-.07	-.60**	-.20
	RR %	-.00	.09	.01	.18	.19	.18	-.05	1.00**	-.04	.27	_	.27	.07	.23	.09
	Int %	.30*	.43	.31**	-.01	-.22	-.03	.05	1.00**	.07	-.13	_	-.13	.05	.23	.06
	Total sel %	.16	.01	.14	.09	-.44	.02	-.07	-1.00**	-.12	.11	_	.11	.05	-.16	.01
N	100	46	14	60	41	10	51	20	2	22	9	0	9	47	14	61
2009																
	CV %	-.22*	-.06	-.19*	-.14	-.43	-.16	.01	-1.00**	-.06	-.49	-1.00**	-.50*	-.27**	-.13	-.24*
	RR %	.13	-.21	.12	.25*	-.31	.23*	.29	-1.00**	.29	.23	-1.00**	.23	.26*	-.28	.25*
	Int %	.09	.32	.10	.05	.54	.06	.04	1.00**	.03	.16	1.00**	.11	.14	.46	.14
	Total sel %	.00	.09	.02	.10	.05	.09	.29	1.00**	.22	-.21	1.00**	-.24	.10	.11	.11
N	107	67	10	77	58	8	66	24	2	26	10	2	12	68	10	78
2010																
	CV %	.02	.15	.02	-.01	.13	.00	-.00	_	-.00	-.12	_	-.12	-.01	.22	.00
	RR %	.21*	.36	.21**	.09	.22	.11	.09	_	.09	-.01	_	-.01	.18*	.26	.19*
	Int %	-.03	-.00	.00	-.02	.25	-.08	.04	_	.04	-.05	_	-.05	-.02	.11	-.03
	Total sel %	.13	.23	.15*	.03	.29	.02	.08	_	.08	-.13	_	-.13	.10	.29	.11
N	140	100	17	117	74	8	82	34	0	34	12	0	12	101	17	118
Combined years																
	CV %	-.07	-.22	-.09	-.09	-.39*	-.11	-.02	-.99**	-.08	-.22	-1.00**	-.24	-.10	-.33*	-.12*
	RR %	.10	-.01	.09	.09	.22	.11	.05	.07	.05	.21	-1.00**	.21	.13*	.05	.13*
	Int %	.10	.31*	.12*	.02	.07	-.01	.05	.31	.06	-.03	1.00**	-.04	.05	.31*	.05
	Total sel %	.09	.09	.08	.01	-.03	-.01	.06	-.97*	.02	.01	1.00**	-.01	.05	.07	.05
N	347	213	41	254	173	26	199	78	4	82	31	2	33	216	41	257

Notes. CV% = CV score as a percentage. RR% = Referee reports score as a percentage. Int % = Interview score as a percentage. Total sel % = Total selection score as a percentage. MiniCEX % = MiniCEX Assessment score; Average MiniCEX % = Mean score of MiniCEX 1-4
r = Pearson Correlation. N = Number of trainees.
** = Correlation is significant at .01 (1-tailed). * = Correlation is significant at .05 (1-tailed). Teal = Also significant in regression analysis.

Table 66 *Significant correlations between GS ANZ selection and ETA performance for 2008, 2009, 2010 selection years and combined selection years; including n < 5*

		Pearson Correlations (*r*)														
		ETA1 %			ETA2 %			ETA3 %			ETA4 %			Average ETA %		
		Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ	Au	NZ	ANZ
2008																
	CV %	-.04	-.20	-.08	.12	-.32	.03	**-.28***	-.17	**-.25***	-.21	.07	-.14	-.13	-.21	-.15
	RR %	.14	-.18	.08	.11	-.13	.06	.12	.21	.14	**.30***	-.00	**.24***	.16	-.03	.12
	Int %	-.01	.26	.02	**-.27***	.33	**-.21***	.12	**.43***	.15	.18	**.38***	**.20***	-.07	**.49***	-.01
	Total sel %	.05	-.13	.02	-.09	-.15	-.11	.06	.23	.09	**.24***	.23	**.24***	-.01	.07	-.01
N	100	59	19	78	60	19	79	55	19	74	50	19	69	62	19	81
2009																
	CV %	-.03	-.11	-.04	-.03	.29	.02	-.19	-.21	**-.21***	-.21	.24	-.16	-.16	.03	-.14
	RR %	**.29****	.18	**.27****	**.20***	**.55***	**.24***	**.41****	.06	**.37****	**.40****	**.66****	**.42****	**.48****	**.39***	**.46****
	Int %	.12	**.60****	**.19***	.18	.37	**.21***	-.03	**.80***	.07	.04	.16	.07	.09	**.57****	**.19***
	Total sel %	**.29****	.36	**.30****	**.26****	**.65****	**.33****	.16	.55	**.19***	.17	**.57***	**.23***	**.32****	**.55****	**.36****
N	107	77	16	93	75	15	90	68	7	75	56	14	70	79	18	97
2010																
	CV %	.10	-.04	.09	-.04	-.17	-.04	_	_	_	_	_	_	-.02	-.18	-.03
	RR %	**.28****	-.06	**.25****	.01	**.43***	.09	_	_	_	_	_	_	**.19***	.26	**.22****
	Int %	.06	-.01	.01	-.00	.22	-.05	_	_	_	_	_	_	.04	.11	-.04
	Total sel %	**.30****	-.05	**.23****	-.03	.28	-.02	_	_	_	_	_	_	.14	.13	.09
N	140	90	16	106	110	18	128	0	0	0	0	0	0	116	18	134
Combined years																
	CV %	-.02	-.15	-.04	.01	-.05	.01	**-.23****	-.09	**-.23****	**-.17***	.15	-.11	**-.12***	-.06	**-.10***
	RR %	**.16****	-.05	**.13***	.06	.13	.07	**.27****	.07	**.26****	**.33****	.15	**.30****	**.19****	.18	**.19****
	Int %	.05	**.29***	.08	-.08	.21	-.06	.06	**.54****	.12	.10	**.28***	.12	.02	**.24***	.04
	Total sel %	**.13***	.07	**.11***	-.01	.21	.01	.11	.28	.13	**.20***	**.29***	**.23****	.06	**.25***	.08
N	347	226	51	277	245	52	297	123	26	149	106	33	139	257	55	312

Notes. CV% = CV score as a percentage. RR% = Referee reports score as a percentage. Int % = Interview score as a percentage. Total sel % = Total selection score as a percentage. ETA % = End of Term Assessment score; Average ETA % = Mean score of ETAs 1-4
r = Pearson Correlation. N = Number of trainees.
** = Correlation is significant at .01 (1-tailed). * = Correlation is significant at .05 (1-tailed). Teal = Also significant in regression analysis.

Regression analyses – General Surgery – selection items and assessments: Tables

Table 67 *Summary of selection items and examination scores having regression value (ANOVA) sig ≤ .05*

		GenSSE	SpecSSE	CE
2008	**Au**			
	NZ			
	ANZ			
2009	**Au**			**Au CE**◈ $F(3,84) = 6.960$ $p < .001$ $R^2 = .199$
	NZ			
	ANZ		**ANZ SpecSSE**▲ $F(3,89) = 3.082$ $p = .031$ $R^2 = .094$	**ANZ CE**◈ $F(3,101) = 7.300$ $p < .001$ $R^2 = .178$
2010	**Au**			
	NZ	**NZ GenSSE**◈ $F(3,15) = 6.015$ $p = .007$ $R^2 = .546$	**NZ SpecSSE**◈ $F(3,14) = 5.485$ $p = .011$ $R^2 = .540$	**NZ CE**▲ $F(3,15) = 3.377$ $p = .046$ $R^2 = .403$
	ANZ			
Combined years	**Au**			**Au CE**◈ $F(3,279) = 9.471$ $p < .001$ $R^2 = .092$
	NZ	**NZ GenSSE**◈ $F(3,50) = 4.997$ $p = .004$ $R^2 = .194$	**NZ SpecSSE**◈ $F(3,37) = 4.426$ $p = .009$ $R^2 = .264$	
	ANZ			**ANZ CE**◈ $F(3,333) = 11.038$ $p < .001$ $R^2 = .090$

▲ Selection items model p ≤ .05 (Total sel p > .05).
◈ Selection items model p ≤ .05 and Total sel p ≤ .05.
⦿ Total sel p ≤ .05 and one or more selection items p ≤ .05 but Selection items model p > .05.
⬣ Total sel p ≤ .05 and individual selection items p > .05 and Selection items model p > .05. (i.e. selection tools not individually significant).

Table 68 *Summary of selection items and DOPS' scores having regression value (ANOVA) sig ≤ .05*

		DOPS 1	DOPS 2	DOPS 3	DOPS 4	Average DOPS
2008	Au					
	NZ	**NZ DOPS1▲** $F(3,10) = 4.845$ $p = .025$ $R^2 = .592$				**NZ Average DOPS▲** $F(3,10) = 3.887$ $p = .044$ $R^2 = .538$
	ANZ					
2009	Au					**Au Average DOPS⬣** $F(1,64) = 4.937$ $p = .030$ $R^2 = .072$
	NZ					
	ANZ					**ANZ Average DOPS⬣** $F(1,74) = 4.219$ $p = .043$ $R^2 = .054$
2010	Au	**Au DOPS1◈** $F(3,98) = 3.525$ $p = .018$ $R^2 = .097$		**Au DOPS3▲** $F(3,30) = 3.179$ $p = .038$ $R^2 = .241$		
	NZ					
	ANZ	**ANZ DOPS1◈** $F(3,115) = 3.475$ $p = .018$ $R^2 = .083$		**ANZ DOPS3▲** $F(3,30) = 3.179$ $p = .038$ $R^2 = .241$		
Combined years	Au	**Au DOPS1◈** $F(3,207) = 2.615$ $p = .052$ $R^2 = .037$	**Au DOPS2▲** $F(3,173) = 2.945$ $p = .034$ $R^2 = .049$			**Au Average DOPS◈** $F(3,210) = 2.689$ $p = .047$ $R^2 = .037$
	NZ					**NZ Average DOPS** $F(3,37) = 2.859$ $p = .050$ $R^2 = .188$
	ANZ	**ANZ DOPS1⬣** $F(1,250) = 4.400$ $p = .037$ $R^2 = .017$	**ANZ DOPS2▲** $F(3,200) = 3.455$ $p = .017$ $R^2 = .049$			

▲ Selection items model $p \leq .05$ (Total sel $p > .05$).
◈ Selection items model $p \leq .05$ *and* Total sel $p \leq .05$.
⦿ Total sel $p \leq .05$ *and* one or more selection items $p \leq .05$ *but* Selection items model $p > .05$.
⬣ Total sel $p \leq .05$ *and* individual selection items $p > .05$ *and* Selection items model $p > .05$. (i.e. selection tools not individually significant).

Table 69 *Summary of selection items and MiniCEX scores having regression value (ANOVA) sig ≤ .05*

		MINICEX 1	MINICEX 2	MINICEX 3	MINICEX 4	Average MINICEX
2008	Au					
	NZ					
	ANZ					
2009	Au					**Au Average MiniCEX▲** $F(3,64) = 3.052$ $p = .035$ $R^2 = .125$
	NZ					
	ANZ					**ANZ Average MiniCEX▲** $F(3,74) = 3.279$ $p = .026$ $R^2 = .117$
2010	Au					
	NZ					
	ANZ					
Combined years	**Au**					
	NZ					**NZ Average MiniCEX▲** $F(3,37) = 3.778$ $p = .018$ $\mathrm{R}^2 = .235$
	ANZ					**ANZ Average MiniCEX⬣** $F(3,253) = 2.899$ $p = .036$ $\mathrm{R}^2 = .033$

▲ Selection items model $p \leq .05$ (Total sel $p > .05$).

⬣ Total sel $p \leq .05$ *and* individual selection items $p > .05$ *and* Selection items model $p > .05$. (i.e. selection tools not individually significant).

Table 70 *Summary of selection items and ETA scores having regression value (ANOVA) sig ≤ .05*

		ETA 1	ETA 2	ETA 3	ETA 4	Average ETA
2008	Au				**Au ETA4▲** $F(3,46) = 2.968$ $p = .042$ $R^2 = .162$	
	NZ			**NZ ETA3◈** $F(3,15) = 3.360$ $p = .047$ $R^2 = .402$		
	ANZ				**ANZ ETA4◈** $F(3,65) = 2.925$ $p = .040$ $R^2 = .119$	
2009	Au	**Au ETA1◈** $F(3,73) = 2.937$ $p = .039$ $R^2 = .108$	**Au ETA2◉** $F(1,73) = 5.405$ $p = .023$ $R^2 = .069$	**Au ETA3▲** $F(3,64) = 5.325$ $p = .002$ $R^2 = .200$	**Au ETA4▲** $F(3,52) = 4.092$ $p = .011$ $R^2 = .191$	**Au Average ETA◈** $F(3,75) = 8.532$ $p = .004$ $R^2 = .254$
	NZ		**NZ ETA2◉** $F(3,13) = 9.699$ $p = .008$ $R^2 = .427$		**NZ ETA4◉** $F(1,12) = 5.875$ $p = .032$ $R^2 = .329$	**NZ Average ETA◈** $F(3,14) = 3.846$ $p = .034$ $R^2 = .452$
	ANZ	**ANZ ETA1◈** $F(3,89) = 3.996$ $p = .010$ $R^2 = .119$	**ANZ ETA2◈** $F(3,86) = 4.181$ $p = .008$ $R^2 = .127$	**ANZ ETA3▲** $F(3,71) = 4.862$ $p = .004$ $R^2 = .170$	**ANZ ETA4◈** $F(3,71) = 4.862$ $p = .004$ $R^2 = .170$	**ANZ Average ETA◈** $F(3,93) = 10.818$ $p = .000$ $R^2 = .259$
2010	Au	**Au ETA1◈** $F(3,86) = 3.789$ $p = .013$ $R^2 = .117$				
	NZ					
	ANZ	**ANZ ETA1◈** $F(3,102) = 3.036$ $p = .033$ $R^2 = .082$				
Combined years	Au			**Au ETA3▲** $F(3,119) = 5.404$ $p = .002$ $R^2 = .120$	**Au ETA4◈** $F(3,102) = 5.178$ $p = .002$ $R^2 = .132$	**Au Average ETA▲** $F(3,253) = 5.3764.223$ $p = .006$ $R^2 = .048$
	NZ			**NZ ETA3◈** $F(3,22) = 4.064$ $p = .019$ $R^2 = .357$		**NZ Average ETA◈** $F(3,51) = 2.979$ $p = .040$ $R^2 = .149$
	ANZ			**ANZ ETA3▲** $F(3,145) = 6.045$ $p = .001$ $R^2 = .111$	**ANZ ETA4◈** $F(3,135) = 5.482$ $p = .001$ $R^2 = .109$	**ANZ Average ETA▲** $F(3,308) = 4.754$ $p = .003$ $R^2 = .210$

Notes. ▲ Selection items model $p \leq .05$ (Total sel $p > .05$).
◈ Selection items model $p \leq .05$ *and* Total sel $p \leq .05$.
⦿ Total sel $p \leq .05$ *and* one or more selection items $p \leq .05$ *but* Selection items model $p > .05$.
⬣ Total sel $p \leq .05$ *and* individual selection items $p > .05$ *and* Selection items model $p > .05$. (i.e. selection tools not individually significant).

Regression analyses – General surgery, comparing selection with assessments

Table 71 *Assessments predicted by CV scores*

Positive predictions		
Year	**Country**	**Assessment**
2010	Au	DOPS1
2010	ANZ	DOPS1
Comb. years	NZ	GenSSE
Total		3

Negative predictions		
Year	**Country**	**Assessment**
2008	NZ	DOPS1
2008	NZ	Average DOPS
2010	NZ	CE
Comb. years	NZ	Average DOPS
Comb. years	ANZ	DOPS2
Comb. years	NZ	Average MiniCEX
Comb. years	Au	ETA3
Comb. years	ANZ	ETA3
Total		8

Table 72 *Assessments predicted by RR scores (all positive)*

Year	Country	Assessment	Year	Country	Assessment
2008	ANZ	ETA4	2010	Au	DOPS1
2008	Au	ETA4	2010	ANZ	DOPS1
2009	Au	CE	2010	ANZ	ETA1
2009	ANZ	CE	2010	Au	ETA1
2009	Au	Average DOPS	Comb. years	NZ	SpecSSE
2009	Au	Average MiniCEX	Comb. years	Au	CE
2009	ANZ	Average MiniCEX	Comb. years	ANZ	CE
2009	ANZ	ETA1	Comb. years	Au	DOPS1
2009	Au	ETA1	Comb. years	Au	Average DOPS
2009	ANZ	ETA2	Comb. years	ANZ	Average MiniCEX
2009	Au	ETA2	Comb. years	Au	ETA3
2009	NZ	ETA2	Comb. years	ANZ	ETA3
2009	ANZ	ETA3	Comb. years	ANZ	ETA4
2009	Au	ETA3	Comb. years	Au	ETA4
2009	ANZ	ETA4	Comb. years	Au	Average ETA
2009	Au	ETA4	Comb. years	NZ	Average ETA
2009	NZ	ETA4	Comb. years	ANZ	Average ETA
2009	Au	Average ETA			
2009	ANZ	Average ETA			
			Total		36

Table 73 *Assessments predicted by Int scores*

Positive predictions			Negative predictions		
Year	**Country**	**Assessment**	**Year**	**Country**	**Assessment**
2008	NZ	ETA3	2010	Au	DOPS3
2009	Au	CE	2010	ANZ	DOPS3
2009	ANZ	CE			
2009	ANZ	SpecSSE			
2009	ANZ	ETA1			
2009	ANZ	ETA2			
2009	NZ	Average ETA			
2009	ANZ	Average ETA			
2010	NZ	GenSSE			
2010	NZ	SpecSSE			
Comb. years	NZ	GenSSE			
Comb. years	NZ	SpecSSE			
Comb. years	Au	CE			
Comb. years	ANZ	CE			
Comb. years	NZ	ETA3			
Comb. years	NZ	Average ETA			
Total		16	Total		2

Note: No 2010 NZ scores available for DOPS3

Table 74 *Assessments predicted by Total sel scores (all positive)*

Year	Country	Assessment
2008	ANZ	ETA4
2009	Au	CE
2009	ANZ	CE
2009	ANZ	SpecSSE
2009	Au	Average DOPS
2009	ANZ	Average DOPS
2009	Au	ETA1
2009	ANZ	ETA1
2009	Au	ETA2
2009	NZ	ETA2
2009	ANZ	ETA2
2009	NZ	ETA4
2009	ANZ	ETA4
2009	Au	Average ETA
2009	NZ	Average ETA
2009	ANZ	Average ETA
2010	NZ	GenSSE
2010	NZ	SpecSSE
2010	Au	DOPS1
2010	ANZ	DOPS1
2010	Au	ETA1
2010	ANZ	ETA1
Comb. years	NZ	GenSSE
Comb. years	NZ	SpecSSE
Comb. years	Au	CE
Comb. years	ANZ	CE
Comb. years	Au	DOPS1
Comb. years	ANZ	DOPS1
Comb. years	Au	Average DOPS
Comb. years	Au	ETA4
Comb. years	ANZ	ETA4
Total		31

APPENDIX B – RACS selection and examination policies and instructions

Policies that applied to the 2008. 2009, and 2010 selection years, and RACS instructions to candidates presenting for the Clinical Examination are presented:

1. RACS Policy: Selection to Surgical Education and Training
2. RACS Policy: Conduct of the Surgical Science Examination – Generic Component
3. RACS Policy: Conduct of the Surgical Science Examination – Specialty Specific component
4. RACS Policy: Conduct of the SET Clinical Examination
5. RACS document: Instructions to candidates presenting for the SET Clinical Examination

Division: Department: Title:	Education and Training Administration Surgical Education and Training Selection to Surgical Education and Training	Ref. No.	ETA_SET_005

1. PURPOSE AND SCOPE

This policy provides guidelines for Specialty Boards for the development their individual policies and procedures for trainee selection to College surgical education and training programs. It has been developed in accordance with the accreditation requirements of the Australian Medical Council (AMC), the Medical Council of New Zealand (MCNZ) and the Brennan Principles. It is expected that each Specialty Board will have its own selection policy which complies with this policy and has additional requirements applicable to the individual specialty.

2. KEYWORDS

Selection; Eligibility; Surgical; Education; Training; Criteria

3. BODY OF POLICY

3.1 General Selection Principles

3.1.1 Selection processes for surgical education and training must comply with AMC and MCNZ accreditation requirements and the Brennan Principles. The Brennan Principles are available on the Australian Government website at www.health.gov.au.

3.1.2 Selection processes must be merit based, free of bias and, to the greatest possible extent, quantifiable.

3.1.3 Selection processes must be compliant with relevant Australian and New Zealand laws and the principles of natural justice and procedural fairness.

3.1.4 Selection processes must be open to external scrutiny and conducted in an accountable manner using documented processes.

3.1.5 Selection processes must be conducted on a national or bi-national basis in Australia and New Zealand.

3.1.6 The opportunity to apply for selection must be publicised in a manner which creates awareness of opportunity for all eligible applicants.

3.2 Eligibility Criteria for Surgical Education and Training

3.2.1 Doctors intending to apply for selection into the SET Program must register via the College website. Doctors not registered cannot lodge an application. Refer to the policy Registration for Selection into Surgical Education and Training.

3.2.2 Applicants must consent to a full criminal history check including the submission of relevant documentation on request to enable this to be undertaken noting that:

5.12.1.1 Where consent is not given by the applicant, they will automatically be deemed ineligible for selection and not considered further in the selection process.

5.12.1.2 Applicants with a relevant criminal conviction will be deemed unsuitable for selection to the training program. A relevant conviction includes, but is not limited to a conviction of a sexual nature, a conviction relating to drug usage and or trafficking, a conviction against liberty, morality and abduction, or a conviction relating to dishonesty, fraud and deception.

5.12.1.3 Failure by an applicant to make full and frank disclosure of their criminal history as requested is grounds to automatically deem the applicant unsuitable for selection, unless the matter is a "spent conviction" under the relevant law.

3.2.3 Applicants must provide documentary evidence at the time of registration for application to surgical training that they have current and valid medical registration from an applicable Medical Board or Council in a jurisdiction in which the training program operates which enables full participation in the training program. Current and valid medical registration is defined as:

a. General (unconditional) registration in Australia.
b. General scope or restricted general scope registration in the relevant specialty2 in New Zealand. Applicants with restricted general scope registration may only apply to the specialty program to which their registration applies.

3.2.4 Applicants to the Australian and New Zealand surgical programs must have permanent residency or citizenship of Australia or New Zealand at the time of registration.

3.2.5 In addition to the generic eligibility criteria, individual training programs may apply additional eligibility requirements applicable to their specialty.

3.3 Selection Criteria and Documentation

3.3.1 Prior to publication each specialty training program must have its selection criteria and process approved by the Education Board. Once approved information is published on the College website. Specialty Boards must ensure that information published on society websites conform with the information on the College website.

3.3.2 Once approved by the Education Board, selection processes and criteria cannot be altered until selection is completed.

3.3.3 The minimum selection criteria and selection tools including any minimum standards or cut-off scores must be declared in the application information. These must be, where possible, objective and quantifiable.

3.3.4 A minimum of three selection tools must be used and these must include a structured curriculum vitae, confidential referee reports or professional performance appraisals and a semi-structured interview.

3.3.5 The application information must include an open declaration of the weighting for each selection tool. The weighting for each selection tool must be within the following ranges:

Curriculum Vitae	15 to 25%
Confidential Referee Report	35 to 45%
Interview	35 to 45%

3.3.6 Where a Board uses an additional selection tool, such as a presentation, this is considered as a form of interview. The combination of the additional selection tool and the interview must fall within the weighting range for the interview.

3.3.7 The semi-structured interview must include a minimum of 2 separate panels. The composition of the interview panel should be proportional to the task, and should include a minimum of two and a maximum of four interviewers

3.3.8 The application information must include an estimate of the number of training positions expected to be offered. Where the number of training positions, or an estimate, is unavailable the number of trainees appointed in the previous year should be made known.

3.3.9 Standardised forms and pre-determined scoring mechanism for selection tools must be used with a view to achieving, where possible, objectivity, comparability and quantification.

3.4 Ranking and Outcome

3.4.1 The total score must be used to rank the applicants. Where the scores of two or more suitable applicants are statistically indistinguishable other factors may be considered in the allocation of training positions.

3.4.2 Applicants may be classified as one of the following:

a. <u>Successful</u> being an eligible applicant who satisfied the minimum selection criteria and for whom a position is available as a consequence of their position on the ranking list.

b. <u>Unsuccessful</u> being an eligible applicant who satisfied the minimum selection criteria but whose position on the ranking list falls outside the number of available positions.

c. <u>Unsuitable</u> being an eligible applicant who failed to satisfy the minimum selection criteria as defined in the application information.

3.4.3 Each Specialty Board selection intake must be such that all trainees selected have the opportunity to progress seamlessly from SET1 to SET2. All available SET2+ training positions for which suitable applicants have been identified must be filled.

3.5 Time Lines, Notification and Feedback

3.5.1 The opening and closing dates for applications will be published on the College website by December of the year prior to selection commencing.

3.5.2 Notification letters must be sent to all applicants on the common announcement date approved by the Education Board.

3.5.3 Notification to unsuccessful and unsuitable applicants must include the applicant's overall standing in the selection process and performance feedback for each selection tool which should include the applicant's overall or decile score. Where applicable, unsuccessful applicants must receive information on the waiting list process.

3.5.4 Those applicants deemed unsuitable or unsuccessful must receive information on the minimum standard they failed to achieve, the process available to seek more detailed feedback and direction to the College Appeals Mechanism.

3.5.5 Applicants must accept or decline the offer of a training position within 10 working days of the common announcement date by completing the appropriate documentation of the specialty training program, including signing the Trainee Agreement.

3.5.6 It is the responsibility of successful applicant to seek and fulfil employment requirements at the hospital where the allocated training position is located.

3.5.7 Where an applicant is already a trainee of another training program, and who is not eligible for deferral to complete the current program, acceptance of an offer will result in their automatic withdrawal from their current training program effective from the commencement date of the new training program.

3.5.8 Subsequent offers to those who are next on the ranked waiting list of unsuccessful applicants for whom a training position becomes available must receive formal notification within 10 working days of the closing date for acceptances of the previous round of offers.

3.5.9 The College must be notified of the outcome of the selection process within two days of the selection committee meeting. The College must be notified of any subsequent changes to the outcome of the selection process within two weeks of the confirmed change.

3.6 Documentation Retention

3.6.1 There should be adequate documentation to enable external scrutiny, audit and evaluation of the selection process. It should enable accurate reconstruction of the original detail and process.

3.6.2 Such documentation must be retained in accordance with the College's Document Retention Schedule.

3.7 Review and Evaluation

3.7.1 There should be an annual formal, regular inclusive review of the selection process which should include the frequency and content of complaints, appeals and their outcomes.

3.7.2 There should be a formal evaluation of the selection process including variables such as the completion rate, attrition rate and dismissal rate.

3.7.3 Compliance with this policy will be reviewed on an annual basis using an Audit Tool developed by the College.

3.8 Appeal

Decisions relating to selection may be reviewed or appealed in accordance with the College Appeals Mechanism.

4. ASSOCIATED DOCUMENTS

Registration for Selection into SET
Policy Trainee Registration and Variation Policy

Approver	CEO
Authoriser	Council

Division: Department: Subject:	Education Development and Assessment Department of Examinations Conduct of the Surgical Science Examination – Generic Component	Ref. No.	EDA_EXA_008

1. PURPOSE AND SCOPE

The Surgical Science Examination is a summative evaluation of the candidates' knowledge, understanding and application of anatomy, physiology and pathology in health and disease. The examination has two distinct components – a generic examination (two papers) and a specialty specific examination (one paper) which may be taken concurrently or independently.

The specialty specific component does not apply to trainees in the Orthopaedic Surgery, Paediatric Surgery or Plastic and Reconstructive Surgery training programs as they are required to sit the relevant examination specific to the corresponding training program.

All Surgical Education and Training (SET) Trainees, regardless of their specialty, are required to undertake the Generic SSE.

The curriculum for the SSE is published on the website at www.surgeons.org.

2. KEYWORDS

Surgical Science Generic Examination, Examinations, Surgical Education and Training. Eligibility, Application Process, Number of Attempts and Time Limitations

3. BODY OF POLICY

3.1 Surgical Science Examination – Generic Component

3.1.1 Format of the Examination

The examination consists of two papers each containing 120 questions. Anatomy, Physiology and Pathology questions are equally represented. The examination is conducted over two consecutive days.

3.1.2 Eligibility

Active and Interrupted Trainees on the SET and Basic Surgical Training (BST) programs are eligible to present for this examination. Applicants who have accepted a place onto a specialty training program but have not yet commenced training, and have not been approved for deferral, are also eligible to apply.

Trainees who have deferred commencement of training are not eligible to present for this examination.

3.1.3 Prerequisites

There are no prerequisites.

3.1.4 Application Process

Candidates are required to submit an application form and pay the SSE examination fee prior to sitting the examination.

Applications will only be accepted up to the advertised closing date. No late applications will be accepted.

The examination date and application closing date are published on the College Examinations' website.

3.1.5 Timeframe

The examination is available twice annually.

3.1.6 Number of Attempts and Time Limitations

Trainees will be dismissed from the training program if they fail to pass the examination after four (4) attempts, or have undertaken two (2) years of active training in the SET program of any specialty, without success in this examination.

3.1.7 Trainees who Transfer to Another Specialty

Should a Trainee transfer to another specialty:

- a successful pass in the Generic SSE examination will be recognised
- transfer to another specialty will not alter the total number of attempts and years of training as outlined under 3.1.6 .

3.1.8 Marking System

3.1.9 Pass Standard

A pass standard is set for each examination based on a standard setting evaluation and on the difficulty of the questions in each examination. A minimum standard is required in each of the anatomy, physiology and pathology components at the same attempt in order to achieve an overall pass mark.

3.1.10 Minimum Standard Pass

A candidate who fails to achieve a pass standard, but who in two separate examinations, achieves a minimum standard in each of the disciplines of anatomy, physiology and pathology, will be deemed to have passed the examination if their score for the whole examination reaches a determined minimum standard on each occasion. The determined minimum standard for the whole examination will be calculated using the same principles as 3.1.9 above.

3.1.11 Component Failure

A candidate cannot carry forward a pass in a component i.e., anatomy, physiology or pathology from one Generic SSE to another examination.

3.1.12 Results

Candidates will be awarded a result of pass or fail only for the Examination. There is no rank order.

Paper 1 and Paper 2 are considered in conjunction to determine the final examination result.

Candidates are formally advised in writing of their result and a breakdown of results can be downloaded from the College website. It is the responsibility of the Examinations Department to communicate results to Specialty Boards.

3.2 Withdrawals

Applicants withdrawing from the examination must notify the Examinations Department of the withdrawal in writing. Applicants who do not formally notify the Examinations Department of their withdrawal will forfeit the examination fee. The following withdrawal fees apply.

Applicants who:

- withdraw from the examination **prior** to the examination closing date will be refunded **100%** of the examination fee.
- withdraw from the examination **more than 10** working days prior to the first scheduled day of the examination will be refunded **50%** of the examination fee.

- withdraw from the examination **10 working days** or less, of the first scheduled day of the examination will be refunded **25%** of the examination fee.

4. ASSOCIATED DOCUMENTS

SSE Generic Examination Application Form
Conduct of the Surgical Science Examination – Specialty Specific Component Policy SET: Notification of Special Circumstances and Disability Policy
Special Circumstances Application Form

Approver Director
Authoriser Council

Division: **Department:** **Subject:**	**Education Development and Assessment Department of Examinations** **Conduct of the Surgical Science Examination – Specialty Specific Component**	**Ref. No.**	EDA_EXA_009

1. PURPOSE AND SCOPE

This policy provides the criteria for the conduct of the Specialty Specific Surgical Science Examination (SSE) for General Surgery, Urology, Neurosurgery, Cardiothoracic Surgery, Vascular Surgery and Otolaryngology Head & Neck Surgery.

For Paediatric Surgery, Orthopaedic Surgery and Plastic and Reconstructive Surgery, refer to the College Examinations' webpage for information on the relevant examinations specific to these specialties.

The Surgical Science Examination is a summative evaluation of the candidates' knowledge, understanding and application of anatomy, physiology and pathology in health and disease. The examination has two distinct components – a generic examination (two papers) and a specialty specific examination (one paper) which may be taken concurrently or independently at a later date.

General Surgery and Urology use a common Specialty SSE paper. Cardiothoracic Surgery, Neurosurgery, Otolaryngology Head & Neck Surgery and Vascular Surgery have individual papers.

The Specialty Specific Surgical Science Examination does not apply to Trainees in the Orthopaedic Surgery, Paediatric Surgery, or Plastic and Reconstructive Surgery training programs, as they are required to sit the relevant examinations specific to the corresponding training program.

2. KEYWORDS

Surgical Science Specialty Specific Examination, Examinations, Surgical Education and Training.

3. BODY OF POLICY

3.1 Format of the Examination

The examination consists of one paper containing 120 questions. Anatomy, physiology and pathology questions are represented relevant to the specialty practice and weighted by the individual specialty. Each Board maintains currency of the relevant Specialty SSE curriculum.

The examination is conducted over one day and is concurrent with and subsequent to the examination days for the Generic SSE.

3.2 Eligibility and Prerequisite

3.2.1 Eligibility

Active and interrupted Trainees of the relevant SET specialty training program are eligible to present for this examination. Applicants who have accepted a place onto a specialty training program but have not yet commenced training, and have not been approved for deferral, are also eligible to apply.

Trainees who have deferred commencement of training are not eligible to present for this examination.

3.2.2 Prerequisites

There are no prerequisites to present for this examination.

3.3 Application Process

Candidates are required to register and pay the SSE examination fee prior to sitting the examination.

Applications will only be accepted up to the advertised closing date. No late applications will be accepted.

The examination date and application closing date are published on the College Examinations' website.

3.4 Timeframe

The examination is available twice annually.

The Specialty SSE can be taken at the same time as the Generic SSE paper, or may be taken at a subsequent sitting.

3.5 Number of Attempts and Time Limitations

Trainees will be dismissed from the training program if they fail to pass the examination after four (4) attempts, or have undertaken two (2) years of active training in the SET program of the relevant specialty, without success in this examination.

3.6 Trainees who Transfer to Another Specialty

Should a Trainee transfer to another specialty:

3.6.1 a successful pass in the Generic SSE examination will be recognised

3.6.2 transfer to another specialty will not alter the total number of attempts and years of training as outlined under 3.5

3.6.3 the Trainee will be required to successfully pass the relevant Specialty SSE within the stated limitations as outlined in 3.5

3.6.4 Trainees transferring between General Surgery and Urology will not be required to resit the Specialty SSE, as the paper is common to both specialties.

3.7 Results

Candidates will be awarded a result of pass or fail only for this Examination. There is no rank order.

Candidates are formally advised in writing of their result and a breakdown of results can be downloaded from the College website. It is the responsibility of the Examinations Department to communicate results to Specialty Boards.

3.8 Withdrawals

Applicants withdrawing from the examination must formally notify the Examinations Department of the withdrawal, in writing. Applicants who do not formally notify the Examinations Department of their withdrawal will forfeit the examination fee. The following withdrawal fees apply:

Applicants who:

5 withdraw from the examination **prior** to the examination closing date will be refunded
100% of the examination fee.

6 withdraw from the examination **more than 10** working days prior to the first scheduled day of the examination will be refunded **50%** of the examination fee.

7 withdraw from the examination **10 working days** or less, of the first scheduled day of the examination will be refunded **25%** of the examination fee.

4. ASSOCIATED DOCUMENTS

SSE Specialty Specific Examination Application form

Conduct of the Surgical Science Examination – Generic Component Policy SET: Notification of Special Circumstances and Disability Policy

Special Circumstances Application Form

Approver Director
Authoriser Council

Division: Department: Subject:	Education Development and Assessment Department of Examinations Conduct of the SET Clinical Examination	Ref. No.	EDA_EXA_004

1. PURPOSE AND SCOPE

The policy provides the criteria for the conduct of the Surgical Education and Training (SET) Clinical Examination (CE).

The emphasis of the clinical examination is on the application of basic science knowledge and understanding and clinical practice relevant to all forms of surgery. The curriculum for the set clinical examination is on the college website at www.surgeons.org. All set trainees, regardless of their specialty must successfully complete the clinical examination.

2. KEYWORDS

Clinical Examination, Examinations, Surgical Education and Training, OSCE.

3. BODY OF POLICY

3.1 SET Clinical Examination

3.1.1 Format of the Examination

The examination consists of 16 Objective Structured Clinical Examination (OSCE) stations made up of the following four (4) types of questions:

- Examination
- Counseling
- Procedure
- History Taking

Examples of stations include patient history taking and examination, demonstration of practical technical skills, the application of basic science knowledge, data acquisition and analysis, counseling and communication skills.

The examination is conducted in teaching hospitals running concurrently, in listed Australian States and New Zealand, provided that there is a minimum of 10 candidates registered for each site, with one Examiner per station. Due to the resources required to appropriately conduct the examination, registrations may be limited to a maximum of 20 candidates per centre.

3.1.2 Eligibility and Prerequisites

a. Eligibility

Active and interrupted Trainees on the SET and Basic Surgical Training (BST) programs are eligible to present for this examination. Applicants who have accepted a place onto a specialty training program but have not yet commenced training, and have not been approved for deferral, are also eligible to apply.

Trainees who have deferred commencement of training are not eligible to present for this examination.

b. Prerequisites

There are no prerequisites.

3.1.3 Application Process

Candidates are required to register and pay the Clinical examination fee prior to sitting the examination.

Applications will only be accepted up to the advertised closing date. No late applications will be accepted.

The examination dates and application closing dates are published on the College Examinations' website.

3.1.4 Timeframe

The examination is available twice annually.

3.1.5 Number of Attempts and Time Limitations

Trainees will cease training if they fail to pass the examination after four (4) attempts, or they have undertaken two (2) years of training without success at the examination.

3.1.6 Trainees who Transfer to Another Specialty

The examination is generic to all specialties and is a requirement for all Trainees to successfully undertake. Both a successful attempt and the number of attempts will carry over should a Trainee transfer between specialty programs.

3.1.7 Marking System

The overall pass mark will be determined from the station pass marks and the standard error of measurement for this examination. Candidates will be scored at each station using a 25 point checklist and a 6 point global rating scale. The pass mark for each station will be determined from the scores of those candidates rated as 'Borderline Pass' or 'Borderline Fail' on the global scale.

3.1.8 Results

Candidates will be awarded a result of pass or fail only for the examination. There is no rank order.

The Examinations Department formally advises candidates in writing of their result. Results of this Examination are also communicated directly to the relevant Specialty Board.

3.1.9 Withdrawals

Applicants withdrawing from the examination must formally notify the Examinations Department of the withdrawal, in writing. Applicants who do not formally notify the Examinations Department of their withdrawal will forfeit the examination fee. The following withdrawal fees apply.

Applicants who:

a. withdraw from the examination **prior** to the examination closing date will be refunded
100% of the examination fee.

b. withdraw from the examination **more than 10** working days prior to the first scheduled day of the examination will be refunded **50%** of the examination fee.

C. withdraw from the examination 10 working days or less, of the first scheduled day of the examination will be refunded 25% of the examination fee.

4. ASSOCIATED DOCUMENTS

Clinical Examination Application form

SET: Notification of Special Circumstances and Disability Policy Special Circumstances Application Form

Approver Director
Authoriser Council

INSTRUCTIONS TO CANDIDATES PRESENTING FOR THE CLINICAL EXAMINATION

Structure, terminology and marking used in the clinical examination.

There are 4 types of assessed clinical stations – two are primarily "communication" stations (History taking and counselling) and two are "hands on" stations (Examination and procedure). Each station will usually be confined to one of the four types and there will be 4 stations of each type in the examination making a total of 16 stations. If you are required to answer one or two specific questions from the examiner after completing the task, you will be told this in the written material at the beginning of the station.

You should extend the normal courtesy to the surrogate by introducing yourself and where appropriate explain what you need to do. Some questions will set clear limits for the candidate (eg. "Do not examine for…" or "Assume that.."). No marks will be awarded for performance outside the stated boundaries of the question.

For each question, the examiner has a check list of the components of the task which provide 75% of the marks for the station. In addition, there is a "global competency score" which makes up the remaining 25% of the marks. The competency scoring for each types of clinical station is shown in the tables after the section below.

Marking system to commence in October 2008

The mark required to pass each station will be the mean score of those candidates judged to have completed the station at a 'borderline pass' or 'borderline fail' level, according to their scores on the Global Rating Scale. The minimum passing score for the whole exam will be the sum of the sixteen station pass marks plus one standard error of measurement. In order to pass the exam, candidates must pass at least two stations of each type (Examination, Communication, History and Procedure) as well as achieving the minimum passing score for the whole exam.

1. History taking

"Take a history…" – The candidate will be required to interview a surrogate patient. You will be given basic information about the presentation of a clinical problem and are required to take a formal structured and relevant clinical history from the surrogate patient who has been provided with basic responses to the appropriate questions. Appropriate questions asked by the candidate will be marked on the examiners checklist. The marks awarded are NOT dependent on the answers given by the surrogate

Overall Competency Score (History)

To be completed by the examiner - **circle ONE number** only

Examples of global assessment criteria	Category	
No rapport, no logical organization, rude, incomplete history, inappropriate offensive questions, no introduction to patient	**Fail**	**1**
	Borderline Fail	**2**
	Borderline Pass	**3**
Good rapport, reasonably logical, considerate, completes most of the history, mostly appropriate questions, reasonable introduction to the patient	**Adequate Pass**	**4**
	Excellent Pass	**5**
Excellent rapport, logical questioning, very considerate, complete history, appropriate inoffensive questions, good introduction to the patient	**Outstanding Pass**	**6**

2. Communication

There are 5 categories in this type of station.

1. ***"General counselling" -*** The candidate will be given information about a specific condition or the result of an investigation and will be expected to communicate this information to the surrogate patient in simple English terms. They may also be asked to discuss options available to the patient or implications of the information.
2. **"*Obtain Informed Consent"* –** Candidates will be told the specific procedure the patient is to undergo. They should briefly assess the patients understanding of a surgical procedure, then provide a simple explanation of the procedure, discuss the specific risks and benefits of the procedure and expected recovery from it.
3. ***"Breaking bad news"* –** The candidate will be supplied with unpleasant news to convey to the surrogate patient demonstrating not only a knowledge of the subject but also an ability to communicate with empathy.
4. ***"Working in a team"* –** The candidate will be expected to provide information to the surrogate who is acting as a member of the team (nurse, paramedic, colleague) providing appropriate information for the clinical scenario presented.
5. ***"Postoperative information"* –** Information about the postoperative status (a complication, or new scenario) of a patient will be provided and the candidate asked to explain the problem and its implications to the surrogate patient.

Simple English terms should be used to convey the information and, where appropriate, pen and paper may be used to provide a diagram for the patient. It is also important to allow the patient an opportunity to ask questions to make sure they understand. The surrogate may have specific questions to ask during the station.

b) Overall Competency Score (communication)

To be completed by the examiner - **circle ONE number** only

Examples of global assessment criteria	**Category**	
No rapport, poor explanation, no empathy, inappropriate language, avoids questions, rude, patronizing answers, no checking of understanding	**Fail**	**1**
	Borderline Fail	**2**
	Borderline Pass	**3**
Good rapport, reasonable explanation, some empathy, attempts to answer questions, reasonably courteous, appropriate answers, little use of technical language, some checking of understanding	**Adequate Pass**	**4**
	Excellent Pass	**5**
Excellent rapport, clear concise explanation, empathetic, highly appropriate answers, courteous, responds in plain English, ensures information is understood	**Outstanding Pass**	**6**

3. Physical Examination

"Perform a relevant physical examination..." - Candidates should perform a physical examination specific to the area of interest. Usually the region will be defined by the question. (e.g. "Examine the right hip and other relevant features of the right lower limb". The qualifier here serves to indicate that in addition to a hip examination, leg length and a brief neurovascular examination should be performed.) Occasionally you may be asked about the examination for a specific condition which may involve many regions. You will need to explain to the examiner what you are doing and why as you proceed.

You will NOT be required to examine genital or female breast regions but if you think it could be relevant to the condition, you should tell the examiner that you would examine the region without doing so. You are NOT required to take a clinical history and ongoing dialogue with the surrogate should be confined to the physical examination (e.g. "does it hurt", "please open your mouth").

Overall Competency Score (Examination)

To be completed by the examiner - **circle ONE number only**

Examples of global assessment criteria	**Category**	
No rapport, no organization, rough, incomplete examination, rude, inappropriate examination, no explanation to patient, embarrasses patient	**Fail**	**1**
	Borderline Fail	**2**
	Borderline Pass	**3**
Good rapport, reasonably well organized, considerate in most of the examination, reasonably courteous, appropriate examination, reasonable explanations given, caring attitude to patient	**Adequate Pass**	**4**
	Excellent Pass	**5**
Excellent rapport, exceptional organization, gentle, complete examination, courteous, appropriate examination, explains clearly to the patient, very caring attitude to patient	**Outstanding Pass**	**6**

4. Procedure

There are three categories of procedure you may be required to demonstrate.

Diagnostic - Procedures for diagnosis or treatment.

Emergency - Trauma and emergency A&E procedures.

Theatre - Procedures routinely performed in operating theatre settings including the safe handling of instruments.

The examiner at these stations will expect the candidate to demonstrate the method of performing the specified procedure on a surrogate patient, a mannequin or a model. A running commentary is expected from the candidate describing the anatomical landmarks, the presence or absence of obvious complicating factors (e.g. previous surgical scar), and a complete step by step description of the procedure using the equipment provided. If equipment you need is not on display you should indicate to the examiner what you would use, how you would use it and why. Invasive procedure should not be performed on surrogate patients but should usually be completed on mannequins or models. As well as a description of your actions to the examiner, some explanation to and interaction with the surrogate is expected. In some questions you may be asked to do a number of short simple procedures.

Overall Competency Score (procedure)
To be completed by the examiner - **circle ONE number** only

Examples of global assessment criteria	Category	
Incompetent, no technique, no dexterity, no sequential organization, task not completed	Fail	1
	Borderline Fail	2
	Borderline Pass	3
Reasonable competence, good technique, reasonable dexterity, good organization of the task, task completed	Adequate Pass	4
	Excellent Pass	5
Highly competent, outstanding technique, outstanding dexterity, excellent sequential organization of the task, task completed	Outstanding Pass	6

APPENDIX C – General Surgery selection regulations

General Surgery Selection Regulations for 2008, 2009 and 2010 selection years are presented:

1. General Surgery Selection Regulations 2008
2. General Surgery Selection Regulations 2009
3. General Surgery Selection Regulations 2010

SELECTION TO SURGICAL EDUCATION AND TRAINING IN GENERAL SURGERY REGULATIONS

For general instructions and guidelines for selection into the SET Program, please refer to the **SET: Selection to Surgical Education and Training Policy** located on the College Website at College Polices.

For information about PreSET, please refer to the **SET: Preparation for Surgical Training (PreSET)** Policy located on the College website at College Policies.

1. INTRODUCTION

1.1 Definition of terms for the purpose of these Regulations

1.1.1 **Applicant** means a person who has submitted an applicant for the Surgical Education and Training Program in General Surgery to the Royal Australasian College of Surgeons.

1.1.2 **Assessor** means the person identified in accordance with these Regulations to evaluate professionally the applicant's performance using the Hospital Assessment Report.

1.1.3 **Board means** the Royal Australasian College of Surgeons.

1.1.4 **Business Days means** Monday to Friday excluding Public Holidays.

1.1.5 **College** or **RACS means** the Royal Australasian College of Surgeons.

1.1.6 **Interview** means the Board in General Surgery semi-structured General Surgery **interview** conducted as part of the selection process.

1.1.7 **Police Report** means a report on the criminal record of a person.

1.1.8 **Referee** means a person identified in accordance with these regulations to evaluate professionally the applicant's performance.

1.1.9 **Relevant Police Force** means any or all of Australian Federal Police and the various State and **Territory** Police Forces and the New Zealand Police Force.

1.1.10 **SET Program** means the Surgical Education and Training Program in General Surgery as approved by the Board in General Surgery.

1.2 Purpose of these Regulations

The purpose of these Regulations is to set forth and establish the principles, terms and conditions of the selection process for the Royal Australasian College of Surgeons Surgical Education and Training (SET) Program in General Surgery for the 2009 intake. This document is a public document.

1.3 Administration and Ownership

1.3.1 The College is the body accredited and authorised to conduct surgical education and training in Australia and New Zealand and in some regions of Asia.

1.3.2 The Board in General Surgery is responsible for the delivery of the Surgical Education and Training Program in General Surgery, the accreditation of hospital posts, and the assessment and supervision of General Surgical Trainees.

1.3.3 The Board in General Surgery delivers the SET program in Australia and New Zealand, and these Regulations apply to both countries.

1.3.4 For further information, refer to the **SET Specialty Boards and Regional Subcommittees Terms of Reference** located on the College Website at College Policies.

2 APPLICATION

2.1 Applicants wishing to apply to the SET Program must first submit a completed Registration Form to the College via the College website.

2.2 Applicants are required to confirm for themselves that they meet the minimum eligibility criteria required by the Board before submitting their completed Registration Form.

2.3 For further information regarding Registration, including fees, please refer to the SET: Registration for Selection into the Surgical Education and Training (SET) Policy available on the College website at College Policies**.**

2.4 Applicants to the SET Program in General Surgery must be one of the following:

a. a current College registered Pre-SET
b. a current College registered Basic Surgical Trainee
c. a current College registered Specialist Surgical Trainee in any specialty
d. a Fellow of the College
e. an International Medical Graduate who has a current specialist assessment from the College

2.5 Applicants must consent to a full criminal history check including the submission of relevant documentation on request to enable this to be undertaken noting that:

a. Where consent is not given by the applicant, they will automatically be deemed ineligible for selection and not considered further in the selection process.
b. Applicants with a relevant criminal conviction will be deemed unsuitable for selection to the training program. A relevant conviction includes, but is not limited to, a conviction of a sexual nature, a conviction relating to drug usage and/or trafficking, a conviction against liberty, morality and abduction, or a conviction relating to dishonesty, fraud and deception.
c. Failure by an applicant to make full and frank disclosure of their criminal history as requested is grounds to automatically deem the applicant unsuitable for selection, unless the matter is a "spent conviction" under the relevant law.

2.6 Applicants must provide documentary evidence of the following at the time of registration for application to surgical training:

a. General (unconditional) registration in Australia.
b. General scope registration in New Zealand.
c. Permanent residency or citizenship of Australia or New Zealand.

2.7 Applications can only be submitted via the College online application system at www.surgeons.org. No other form of application will be accepted and no extensions will be granted. It is the applicant's responsibility to ensure that they allow enough time to complete the application.

2.8 Separate applications must be made for the General Surgery Training Program in Australia and the General Surgery Training Program in New Zealand.

2.9 Applicants to the General Surgery Training Program in Australia have the option of indicating their preferences for the following Regions:

a. New South Wales/Australian Capital Territory
b. Queensland
c. South Australia
d. Victoria/Tasmania
e. Western Australia

2.10 Applicants should number each Region in order of preference according to the following guidelines:

2.10.1 Applicants to the General Surgery Program in Australia may only list preferences for Regions within Australia, and not New Zealand.
2.10.2 Should an applicant not wish to be considered for a post in a particular Region, they should select the "No Preference" option rather than numbering that Region. This will ensure that applicants are not offered positions which they have no desire to accept.
2.10.3 If an applicant wishes to be considered for a post in any region, and is willing to accept a post in any Region offered to them, they should number each Region in order of preference.
2.10.4 Where a position in a particular Region becomes available and the next ranked applicant has not listed that Region as a preference, the position will be offered to the next eligible applicant who has listed that Region as a preference.

3 SELECTION PROCESS OVERVIEW

3.1 Applicants who satisfy the eligibility and application requirements outlined in Section 2 of these Regulations will be considered in open competition for selection to the SET Program.

3.2 On completion of the relevant components of the selection process eligible applicants will be classified as one of the following:

3.2.1 **Unsuitable** being an eligible applicant who failed to satisfy a minimum standard for selection.
3.2.2 **Unsuccessful** being an eligible applicant who satisfied the minimum standards for selection and therefore suitable but who did not rank highly enough in comparison to the intake to be made an offer of a position.

3.2.3 **Successful** being an eligible applicant who satisfied the minimum standards for selection who is therefore suitable and who has ranked highly enough in comparison to the appropriate intake to be made an offer of selection.

3.3 Applicants **must** satisfy two minimum standards to be deemed suitable for selection to the SET Program. The two minimum standards for selection are:

3.3.1 Applicants must receive no less than 4 satisfactory reports from the 3 Reports and 2 Hospital Assessment Reports obtained.

3.3.2 Applicants must receive a score of "Fair" (2) or above for all questions at the Interview.

3.4 Failure to achieve **any one** of the minimum standards will result in the applicant being deemed **unsuitable** and eliminated from the selection process. Applicants who are deemed **unsuitable** will be notified as outlined in clause 8 of these regulations.

3.5 Applicants who satisfy the two minimum standards for selection and the eligibility conditions will be deemed suitable for selection and will be ranked. The ranking will be determined by applying the following weightings to the percentage adjusted score out of 100 obtained for each of the four selection tools, providing an overall percentage score:

3.5.1 Structured Curriculum Vitae 20%
3.5.2 Structured Referee Reports 30%
3.5.3 Structured Hospital Assessment Reports 10%
3.5.4 Semi-structured Interview 40%

3.6 Successful applicants will be offered positions based on their national ranking and Regional preferences.

3.7 It is expected that due to attrition and requests for interruption, there will be several rounds of offers to the SET Program in General Surgery.

3.8 Applicants who have been deemed suitable but who are not ranked highly enough to receive a first round offer to the SET Program will still be considered eligible for subsequent rounds of offers made by the Board.

3.9 Applicants who do not wish to receive a later round offer to the SET Program in General Surgery must advise the Board by the stipulated deadline.

3.10 Applicants who receive an offer to a Region other than their first preference and who wish to be considered eligible for subsequent rounds of offers must indicate this in writing at the time of acceptance. If the applicant fails to do so, the acceptance of the original offer will stand and the applicant will not be considered for subsequent rounds of offers.

3.11 Once an offer has been accepted, the relevant Regional Subcommittee will allocate the successful applicant to a Training Rotation according to the following guidelines:

3.11.1 Allocation will be based on national rank and preference, entry level eligibility (SET 1 or SET 2) and the number of available positions.

3.11.2 While every effort will be made to match applicants to their preference, due to the number posts available this will not always be possible and new trainees are required to accept the rotation allocated to them.

3.11.3 In the interests of fairness, allocations to posts may not be made until after several rounds of offers have been finalised.

3.12 Applicants who have been deemed suitable but who are not ranked highly enough to receive an offer by the final round will be considered **unsuccessful.** Unsuccessful applicants will be notified in writing as outlined in clause 9 of these Regulations.

4 STRUCTURED CURRICULUM VITAE

4.1 The Structured Curriculum Vitae (online application form) captures information relevant to the eligibility of the applicant and the administration of the selection process, in addition to information on experience, education, research, publications, presentations, development activities and referees.

4.2 Each Structured Curriculum Vitae will be scored by the 2 people nominated by the Board without reference to the opinions of others using a structured scoring system. Where any discrepancy occurs in the scores provided by the two scorers the Board Chairman will score the Structured Curriculum Vitae to identify the anomaly and determine the correct score.

4.3 The Structured Curriculum Vitae will be scored out of a potential 25 points. The components scored are:

4.3.1 Surgical and Medical Experience (Maximum 8 points)
4.3.2 Skills Courses (Maximum 2 points)
4.3.3 Qualifications (Maximum 3 points)
4.3.4 Presentations (Maximum 4 points)
4.3.5 Publications (Maximum 4 points)
4.3.6 Prizes/Awards for Excellence (Maximum 2 points)
4.3.7 Leadership (Maximum 2 points)

4.4 Surgical and medical experience are scored according to the following guidelines, up to a maximum of 8 points:

4.4.1 Scoring will only consider terms undertaken in the last 2 years except where 4.4.2 applies.

4.4.2 Where the applicant has been undertaking research towards a higher degree in a medically related discipline, scoring will consider terms undertaken in the last 4 years.

4.4.3 Terms in Surgery or a related discipline shorter than 5 weeks will not be scored.

4.4.4 Medical terms of less than 10 weeks will not be scored.

4.4.5 Terms planned for after the closing date in the year of application will not be scored.

4.5 Skills Courses are scored according to the following guidelines, up to a maximum of 2 points:

4.5.1 Courses must be must be delivered by a recognised training provider as determined by the Board.
4.5.2 Scoring includes those related to professional development in clinical and technical competencies including ASSET, CCrISP, EMST, CLEAR and Statistics for Surgeons.
4.5.3 Scoring includes courses related to the development of professional competencies such as communication, teamwork and leadership.

4.6 Qualifications are scored according to the following guidelines, up to a maximum of 3 points:

4.6.1 Scoring only includes higher degrees successfully completed at the time of application at a recognised institution as determined by the Board.
4.6.2 Scoring does not include primary medical qualifications including the MBBS or overseas equivalent.
4.6.2 Scoring includes Masters degrees, PhDs and MDs, with extra weighting for Masters by thesis.

4.7 Presentations are scored according to the following guidelines, up to a maximum of 4 points:

4.7.1 Scoring only includes presentations relevant to medicine.
4.7.2 Scoring only includes presentations personally given by the applicant.
4.7.3 Scoring only includes presentations at scientific meetings or conferences subject to abstract selection.
4.7.4 Presentations which have sufficiently similar topics or that have been presented at more than one scientific meeting or conference will only be scored once.
4.7.4 Presentations will be weighted depending on national, local or international level.

4.8 Publications are scored according to the following guidelines, up to a maximum of 4 points:

4.8.1 Scoring only includes publications relevant to medicine.
4.8.2 Scoring only includes publications accepted for publication in a peer reviewed publication and excludes published abstracts.
4.8.3 Each publication can only be scored once.
4.8.4 Scoring includes case reports, articles and book chapters with extra weighting on articles and book chapters where the applicant is the first author.
4.8.5 Documentary evidence of acceptance for publication must be provided at the time of application.

4.9 Prizes will be scored according to the following guidelines, up to a maximum of 2 points:

4.9.1 Scoring only includes prizes or awards for excellence in a medically related field, including prizes for presentations.

4.10 Leadership will be scored according to the following guidelines, up to a maximum of 2 points:

4.10.1 Applicants may score for an elected or appointed position of responsibility on a board, committee or other appropriate body in a community service or professional organisation, as determined by the Board.

4.10.2 Applicants may score for community and cultural involvement or sporting activities as determined by the Board, with weightings for representation at a state, national or international level.

4.11 The score out of 25 will be adjusted to a score out of 20 to correspond to the weighting of the Structured Curriculum Vitae.

5. STRUCTURED REFEREE REPORTS

5.1 References are collected to obtain information, in confidence, about the history of the applicant as well as assessments regarding a number of areas such as personal attributes, quality of work and suitability for the SET Program.

5.2 The applicant must provide contact details including a valid email address for the two (2) supervising consultants who had the greatest period of supervision over the applicant for each rotation undertaken in the two (2) years prior to the closing date for applications; except

5.3 For applicants undertaking research towards a higher degree in a medically related discipline at the time of application, contact details including a valid email address must be provided for the two (2) supervising consultants who had the greatest period of supervision over the applicant for each rotation undertaken in the four (4) years prior to the closing date for applications.

5.4 If an applicant elects not to provide the details for supervising consultants as stipulated by these Regulations, or it is subsequently discovered that the applicant has provided incorrect or misleading information either intentionally or by mistake, including listing supervising consultants who do not strictly comply with these Regulations or omitting supervising consultants in preference for others who have had a lesser supervisory role, the applicant may be automatically withdrawn from the selection process and their application will not considered further in the selection process.

5.5 The units in which the applicant has worked may be contacted as part of the selection process to verify that the supervising consultants listed on the application form comply with these Regulations. The supervising consultants will also be asked to verify compliance with these regulations.

5.6 The Board will select three (3) supervising consultants to be contacted as partof the selection process. In selecting supervising consultants the Board will endeavour to obtain **at least** one report from each General Surgery term (where applicable) and the remaining from other terms with consideration given to the duration and type of term. Reports from supervising consultants for general practice, private assisting or research periods **will not** be sought.

5.7 The Board will select two (2) alternate supervising consultants to be contacted if 5.7 required.

5.8 The supervising consultants selected to submit reports will be at the discretion of the Board and the names **will not** be released to the applicants.

5.9 Where the Board has not obtained three (3) valid reports from supervising consultants identified in accordance with clauses 5.2 and 5.3 within 2 weeks of the reports being issued, the Board will attempt to obtain referee reports from the alternate supervising consultants selected as per clause 5.7.

5.10 Where the Board has requested alternate reports in accordance with clause 5.9, the first three (3) valid reports returned will be used and any reports subsequently received will be discarded.

5.11 If, having applied clauses 5.6 and 5.7, the Board is unable to obtain three (3) valid reports prior to the final submission date, the applicant will be formally withdrawn from the selection process and their application will not be considered further.

5.12 The Board is responsible for the collection of the reports. Applicants will not be provided with updates on the reports collected; nor will they be involved in the collection process in any way. All supervising consultants contacted as part of the selection process will be advised of the confidential nature of the reports. Harassment of any kind of any individual involved in the completion or collection of the reports is a serious matter and may result in the applicant being deemed **unsuitable** for selection and removed from the selection process. Harassment includes repeated requests by the applicant to any supervising consultant for a copy of the report submitted.

5.13 On the report the supervising consultant will be asked to select one of five options for each of the twenty assessment areas which they believe best describes the applicant. The selection criteria which will be scored within the reports can be generally categorised as follows:

5.13.1 Medical Expertise – maximum 8 points
5.13.2 Technical Expertise – maximum 24 points
5.13.3 Judgement/Clinical Decision Making – maximum 24 points
5.13.4 Communication – maximum 16 points
5.13.5 Collaboration – maximum 16 points
5.13.6 Scholar and teacher – maximum 16 points
5.13.7 Professionalism – maximum 56 points

5.14 The options chosen by the referee will be converted to the associated numeric score by the Board using a predetermined scoring system as follows:

5.14.1 The first option is categorised as “unsatisfactory” and scores 0 points.
5.14.2 The second option is categorised as “basic” and scores 2 points.
5.14.3 The third option is categorised as “intermediate” and scores 4 points.
5.14.4 The fourth option is categorised as “advanced” and scores 6 points.

5.14.5 The fifth option is categorised as “expert” and scores 8 points.

5.15 The individual report scores will be converted to a percentage score rounded to two decimal places, calculated by dividing the total score for the report by the total numbers of questions for which the referee has provided a response.

5.16 If the referee has not provided a response for more than four (4) of the questions, the report will be invalid and will not be used as part of the selection process.

5.17 The percentage scores for the three (3) individual reports collected will be combined and expressed as a score out of 30.

5.18 Applicants must receive a score of “Basic” (2) or above for each question identified by the Board as Essential Criteria in order for the report to be deemed satisfactory.

5.19 Applicants who receive two (2) or more reports classified as unsatisfactory in accordance with clause 5.18 out of the three (3) referee reports and two (2) hospital assessment reports collected will be deemed unsuitable and eliminated from the selection process.

6 STRUCTURED HOSPITAL ASSESSMENT REPORTS

6.1 Hospital Assessment Reports are collected to obtain information, in confidence, about the history of the applicant as well as assessments regarding a number of areas such as personal attributes, quality of work and suitability for the SET Program.

6.2 The aim of the Reports is to obtain feedback from the perspective of assessors other than supervising consultants.

6.3 The applicant must provide contact details including a valid email address for two (2) of the following assessors for each rotation undertaken in the last two (2) years (or four (4) years for applicants undertaking research towards a higher degree in a medically related discipline at the time of application):

6.3.1 Hospital Medical Officer (HMO) Manager
6.3.2 Director of Medical Services
6.3.3 Director of Clinical Training
6.3.4 Director of Intern Training
6.3.5 Unit Manager
6.3.6 Allied Health Manager
6.3.7 Medical Superintendent

6.4 If an applicant elects not to provide the details for assessors as stipulated by these Regulations, or it is subsequently discovered that the applicant has provided incorrect or misleading information either intentionally or by mistake, the applicant may be automatically withdrawn from the selection process and their application will not considered further in the selection process.

6.5 The units in which the applicant has worked may be contacted as part of the selection process to verify that the assessors listed on the application form comply with these Regulations. The assessors will also be asked to verify compliance with these regulations.

6.6 The Board will select two (2) assessors to be contacted as part of the selection process. In selecting assessors the Board will endeavour to obtain **at least** one report from each General Surgery term (where applicable) and the remaining from other terms with consideration given to the duration and type of term.

6.7 The Board will select two (2) alternate assessors to be contacted if required.

6.8 The supervising consultants selected to submit reports will be at the discretion of the Board and the names **will not** be released to the applicants.

6.9 Where the Board has not obtained two (2) valid reports from supervising consultants identified in accordance with clauses 6.3 within 2 weeks of the reports being issued, the Board will attempt to obtain referee reports from the alternate supervising consultants selected as per clause 6.7.

6.10 Where the Board has requested alternate reports in accordance with clause 6.7 the first two (2) valid reports returned will be used and any reports subsequently received will be discarded.

6.11 If, having applied clauses 6.3 and 56.9, the Board is unable to obtain two (2) valid reports prior to the final submission date, the applicant will be formally withdrawn from the selection process and their application will not be considered further.

6.12 The Board is responsible for the collection of the reports. Applicants will not be provided with updates on the reports collected; nor will they be involved in the collection process in any way. All assessors contacted as part of the selection process will be advised of the confidential nature of the reports. Harassment of any kind of any individual involved in the completion or collection of the reports is a serious matter and may result in the applicant being deemed **unsuitable** for selection and removed from the selection process. Harassment includes repeated requests by the applicant to any assessor for a copy of the report submitted.

6.13 On the report the assessors will be asked to select one of five options for each of the eleven assessment areas which they believe best describes the applicant. The selection criteria which will be scored within the reports can be generally categorised as follows:

6.13.1 Communication – maximum 16 points
6.13.2 Collaboration – maximum 16 points
6.13.3 Professionalism – maximum 56 points

6.14 The options chosen by the referee will be converted to the associated numeric score by the Board using a predetermined scoring system as follows:

6.14.1 The first option is categorised as “unsatisfactory” and scores 0 points.
6.14.2 The second option is categorised as “basic” and scores 2 points.
6.14.3 The third option is categorised as “intermediate” and scores 4 points.
6.14.4 The fourth option is categorised as “advanced” and scores 6 points.

6.14.5 The fifth option is categorised as “expert” and scores 8 points.

6.15 The individual report scores will be converted to a percentage score rounded to two decimal places, calculated by dividing the total score for the report by the total numbers of questions for which the referee has provided a response.

6.16 If the referee has not provided a response for more than four (4) of the questions, the report will be invalid and will not be used as part of the selection process.

6.17 The percentage scores for the two (2) individual reports collected will be combined and expressed as a score out of 10.

6.18 Applicants must receive a score of “Basic” (2) or above for each question identified by the Board as Essential Criteria in order for the report to be deemed satisfactory.

6.19 Applicants who receive two (2) or more reports classified as unsatisfactory in accordance with clause 5.18 out of the three (3) referee reports and two (2) hospital assessment reports collected will be deemed unsuitable and eliminated from the selection process.

7 SEMI-STRUCTURED INTERVIEW

7.1 The aim of the interview is to obtain information relevant to the suitability of the applicant.

7.2 In Australia, the interviews will be held at the Regional Offices of RACS in New South Wales, Queensland, Victoria, South Australia and Western Australia.

7.3 In New Zealand, the interviews will be held at the Wellington Office of RACS.

7.4 Interview dates will be published on the College website prior to the opening of the application process.

7.5 Applicants who satisfy the minimum standard outlined in clause 3.3.1 of these Regulations will receive an interview time via email at least ten (10) business days prior to the interview date. Applicants who do not satisfy the minimum standard will not be eligible to attend an interview.

7.6 It is the applicant’s responsibility to make the appropriate travel arrangements and to meet any costs incurred in attending the interview. The Board accepts no responsibility for any costs incurred by applicants in attending the interview or applicants who fail to satisfy the minimum standards or eligibility who are not permitted to attend an interview.

7.7 Applicants must make themselves available at the scheduled interview time. Applicants who do not present for the interview at the scheduled time will not be considered further in the selection process and their application will be **withdrawn.**

7.8 In Australia, interviews will be conducted by one panel of two (2) interviewers per 7.1 applicant. Interviews will be approximately 30 minutes in duration.

7.9 In New Zealand interviews will be conducted by four (4) panels each comprising two (2) interviewers. The interviews will be approximately 45 minutes duration.

7.10 The interview will consist of (six) 6 sections designed to assess criteria set out in clause 7.11.as well as several questions relating to regional cultural considerations and may include scenario based questions.

7.11 The interview consists of the following six (6) sections:

General Surgical Insight and Self-Motivation
Ethics
Audit
Patient Care Skills
Team Skills
Overall Interview and Communication Skills

7.12 The Board will approve several questions for each section of the interview. Each section will be assessed at each interview, but specific questions may differ between regions.

7.13 Applicant responses will be assessed against criterion statements developed for each section and scored according to the following scale:

Poor (1) – Significantly better than the criterion statement.
Fair (2) – Response compares exactly with the criterion statement.
Good (3) - Better in some aspects than the criterion statement.
Very (4) – Generally better than the criterion statement.
Excellent (5) – Significantly better than the criterion statement.

7.14 Each interview panel will provide a consensus score for each of the sections they administer.

7.15 Applicants must receive a score of Fair or above for each section in order to achieve the minimum standard for this selection tool.

7.16 Applicants who fail to achieve the minimum standard will be deemed **unsuitable** and will be eliminated from further participation in the selection process.

8 FEEDBACK TO UNSUITABLE APPLICANTS

8.1 Applicants who have been deemed unsuitable for selection will not be considered further in the selection process. These applicants will be notified in writing of the following:

8.1.1 That they have been deemed unsuitable for selection and will not be considered further in the selection process.
8.1.2 Information on the overall scores received for each of the selection tools completed.
8.1.3 Notification of the minimum standard of selection process Regulation which they filed to satisfy.

8.1.3 A breakdown of the scores received for each selection tool employed prior to being eliminated from the selection process. Verbal feedback will not be supplied.

8.1.4 Should applicants desire further feedback, they may discuss the areas in which they were found to be deficient with their supervisors.

9 FEEDBACK TO UNSUCCESSFUL APPLICANTS

9.1 Applicants who have been deemed unsuccessful will be notified in writing of the following:

10.1.1 That they have been deemed suitable for selection but have not ranked highly enough to be made an offer in accordance with the intake and have therefore been unsuccessful.

10.1.2 Information on the overall scores received for each of the selection tools completed and their national ranking.

10.1.3 Information on the waiting list process and their position in the list should a position, in accordance with the intake, become available.

10.1.4 Information on the process available to seek more detailed written feedback.

10 FEEDBACK TO SUCCESSFUL APPLICANTS

10.2 Applicants who have been deemed successful in the selection process will be notified in writing and by email of the following:

10.2.1 That they have been successful in the selection process and are being offered a position on the SET Program subject to the conditions outlined in clause 11.

10.2.2 Information on the overall scores received for each of the selection tools completed and their national ranking.

10.2.3 Information on applicable entry point (SET1 or SET2) and the regional allocation and on the process for allocation to a training post.

10.3 Acceptance of the offer to the SET Program will be conditional on the following:

10.3.1 The applicant being registered in the state/territory of offer

10.3.2 The applicant being employable by the relevant health areas and/or the allocated hospital.

10.3.3 The information submitted in the application form being true and correct.

10.3.4 Satisfactory completion of all minimum eligibility criteria by 31 December in the year of application.

10.3.5 Provision of any outstanding documentation required by the Board.

10.3.6 Payment of all monies owed to the College.

10.4 Applicants who fail to satisfy any of the conditions outlined in clause 10.2 of these Regulations will automatically forfeit the offer.

10.5 Applicants who fail to return the acceptance of offer form by the stipulated deadline, or who decline the offer, will automatically forfeit the offer.

10.6 Applicants who return the acceptance of offer form by the stipulated deadline, and who satisfy the conditions outlined in clause 10.2 will be contacted by the relevant Regional Office in accordance with clause 3.11 of these Regulations.

11 DEFERRAL OF TRAINING

11.1 Applicants who wish to defer the commencement of their General Surgical Education and Training must lodge the request to the Board in General Surgery at the time of acceptance using the following procedure:

11.1.1 Lodge the request including any relevant documentation through the "Trainee Online Request" facility on the College Website.

11.1.2 The request will be considered by the relevant Regional Subcommittee and forwarded to the Board for approval.

11.2 All applications for deferral or interruption are governed by the **SET: Trainee Registration and Variation Policy** available on the College website at Policies and Procedures.

11.3 Applicants are required to ensure that their requests for deferral comply with the above policy.

11.4 The Board in General Surgery does not have the authority to alter College Policy, or to approve non-compliant requests.

12 CONJOINT COMMITTEE FOR THE RECOGNITION OF TRAINING IN GASTROINTESTINAL ENDOSCOPY

12.1 The CCRTGE is a national body comprising representatives from the Gastroenerological Society of Australia, the Royal Australasian College of Physicians and the Royal Australasian College of Surgeons.

12.2 General Surgery Trainees are required to register with the CCRTGE. Upon acceptance of a position on the program, Trainee contact details will be forwarded to the CCRTGE in order to facilitate registration.

13 RURAL SURGICAL TRAINING

13.1 The Rural Surgical Training Program (RSTP) aims to provide high quality, cost effective surgical training in rural Australia. It offers the opportunity for Trainees to undertake a flexible program which can include regional, rural and remote practice. In addition to gaining broad surgical experience, RSTP Trainees have access to a large network of rural surgeons and mentor assistance, as well as financial assistance for conferences and training courses.

13.2 The contact details for Trainees who have indicated their interest in the RSTP will be forwarded to the Rural Surgical Training Board.

SELECTION TO SURGICAL EDUCATION AND TRAINING IN GENERAL SURGERY REGULATIONS

For general instructions and guidelines for selection into the SET Program, please refer to the SET: Selection to Surgical Education and Training Policy located on the College Website.

1. INTRODUCTION

2.7 Definition of terms for the purpose of these Regulations

2.7.1 Applicant means a person who has submitted an application for the Surgical Education and Training Program in General Surgery to the Royal Australasian College of Surgeons.

2.7.2 Board means the Royal Australasian College of Surgeons Board in General Surgery.

2.7.3 Business Days means Monday to Friday excluding Public Holidays.

2.7.4 College or RACS means the Royal Australasian College of Surgeons.

2.7.5 Interview means the Board in General Surgery semi-structured General Surgery panel interview conducted as part of the selection process.

2.7.6 Police Report means a report on the criminal record of a person.

2.7.7 Referee means a person identified in accordance with these Regulations to evaluate professionally the applicant's performance.

2.7.8 Relevant Police Force means any or all of Australian Federal Police and the various State and Territory Police Forces and the New Zealand Police Force.

2.7.9 SET Program means the Surgical Education and Training Program in General Surgery as approved by the Board in General Surgery.

2.8 Purpose of these Regulations

The purpose of these Regulations is to set forth and establish the principles, terms and conditions of the selection process for the Royal Australasian College of Surgeons Surgical Education and Training (SET) Program in General Surgery for the 2010 intake. This is a public document.

2.9 Administration and Ownership

2.9.1 The College is the body accredited and authorised to conduct surgical education and training in Australia and New Zealand and in some regions of Asia.

2.9.2 The Board in General Surgery is responsible for the delivery of the Surgical Education and Training Program in General Surgery, the accreditation of hospital posts, and the assessment and supervision of General Surgical Trainees.

2.9.3 The Board in General Surgery delivers the SET Program in General Surgery in Australia and New Zealand, and these Regulations apply to both countries.

2.9.4 For further information, refer to the SET Specialty Boards and Regional Subcommittees Terms of Reference located on the College Website.

2. APPLICATION

13.1 Applicants wishing to apply to the SET Program in General Surgery must first submit a completed Registration Form to the College via the College website by the published closing date.

13.2 Applicants are required to confirm for themselves that they meet the minimum eligibility criteria required by the Board before submitting their completed Registration Form. Only applicants who satisfy the eligibility and application requirements in accordance with College policy will be considered in open competition for selection to the SET Program in General Surgery.

13.3 For further information regarding Registration, including fees, please refer to the SET: Registration for Selection into the Surgical Education and Training (SET) Policy available on the College Website.

13.4 Applicants must consent to a full criminal history check including the submission of relevant documentation on request to enable this to be undertaken noting that:

1. Where consent is not given by the applicant, they will automatically be deemed ineligible for selection and not considered further in the selection process.
2. . Applicants with a relevant criminal conviction will be deemed unsuitable for selection to the training program. A relevant conviction includes, but is not limited to, a conviction of a sexual nature, a conviction relating to drug usage and/or trafficking, a conviction against liberty, morality and abduction, or a conviction relating to dishonesty, fraud and deception.
3. Failure by an applicant to make full and frank disclosure of their criminal history as requested is grounds to automatically deem the applicant unsuitable for

selection, unless the matter is a "spent conviction" under the relevant law.

13.5 Applicants must have current and valid medical registration from the applicable Medical Board or Council in Australia or New Zealand at the time of registration. Australian applicants must have general (unconditional) registration. New Zealand applicants must have general scope registration without restriction or general scope registration restricted to general surgery.

13.6 Applicants must have citizenship or have been granted permanent residency status in Australia or New Zealand at the time of registration.

13.7 Applications can only be submitted via the College online application system at www .surgeon s.org. No other form of application will be accepted and no extensions will be granted. It is the applicant's responsibility to ensure that they allow enough time to complete the application.

13.8 Separate applications must be made for the SET Program in General Surgery in Australia and the SET Program in General Surgery in New Zealand. Applicants can not apply for both programs.

13.9 Applicants to the SET Program in General Surgery in Australia have the option of indicating their preferences for the following Regions:

2.1.1 New South Wales/Australian Capital Territory
2.1.2 Queensland
2.1.3 South Australia
2.1.4 Victoria/Tasmania
2.1.5 Western Australia

13.10 Applicants should number each Region in order of preference according to the following guidelines:

1.1.1 Applicants to the SET Program in General Surgery in Australia may only list preferences for Regions within Australia, and not New Zealand.

1.1.2 Should an applicant not wish to be considered for a post in a particular Region, they should select the "No Preference" option rather than numbering that Region. This will ensure that applicants are not offered positions that they have no desire to accept.

1.1.3 If an applicant wishes to be considered for a post in any Region, and is willing to accept a post in any Region offered to them, they should number each Region in order of preference.

1.1.4 Where a position in a particular Region becomes available and the next ranked applicant has not listed that Region as a preference, the position will be offered to the next eligible applicant who has listed that Region as a preference.

13.11 Applicants must note the following General Surgery specific eligibility requirement:

Rotation Type	Minimum Duration	Validity Period	Completed By
Surgery in general (Surgery in any surgical discipline / sub-specialty).	2 x 8 week	2 years or 4 years if a period of full time study in a medically related	By the end of 2009
Surgery in critical care. Trauma / ICU / HDU / ED	1X 8 week	As above	By the end of 2009

3. SELECTION PROCESS OVERVIEW

10.1 Applicants who satisfy the eligibility and application requirements in accordance with College policy will be considered in open competition for selection to the SET Program in General Surgery.

10.2 On completion of the relevant components of the selection process, eligible applicants will be classified as one of the following:

2.9.5 Unsuitable being an eligible applicant who failed to satisfy a minimum standard for selection.

2.9.6 Unsuccessful being an eligible applicant who satisfied the minimum standards for selection who is therefore suitable but who did not rank highly enough in comparison to the intake to be made an offer of selection.

2.9.7 Successful being an eligible applicant who satisfied the minimum standards for selection who is therefore suitable and who has ranked highly enough in comparison to the appropriate intake to be made an offer of selection.

10.3 Applicants must satisfy the two (2) minimum standards to be deemed suitable for selection to the SET Program in General Surgery. The two minimum standards for selection are:

4.1.1 Applicants must score an overall percentage adjusted score of 64% or above in the Structured Referee Reports scoring process.

4.1.2 Applicants must score "Basic" (2) or above for all questions at the interview.

10.4 Applicants who satisfy the two (2) minimum standards for selection and the eligibility conditions will be deemed suitable for selection and will be ranked. The ranking will be determined by applying the following weightings to the percentage adjusted score out of 100 obtained for each of the three (3) selection tools, providing an overall percentage score:

3.1.1 Structured Curriculum Vitae 20%
3.1.2 Structured Referee Reports 40%
3.1.3 Semi-Structured General Surgery Panel I nterview 40%

10.5 Successful applicants will be offered positions based on their national ranking and Regional preferences. (There are no regional rankings for the New Zealand program; those applicants will be offered positions on their national ranking only.)

10.6 It is expected that due to attrition and requests for interruption, there will be several rounds of offers to the SET Program in General Surgery.

10.7 Applicants who have been deemed suitable but who are not ranked highly enough to receive a first round offer to the SET Program in General Surgery will still be considered eligible for subsequent rounds of offers made by the Board.

10.8 Applicants who do not wish to receive a later round offer to the SET Program in General Surgery must advise the Board by the stipulated deadline.

10.9 Applicants for the Australian program who receive an offer to a Region other than their first preference and who wish to be considered eligible for subsequent rounds of offers must indicate this in writing at the time of acceptance. f the applicant fails to do so, the acceptance of the original offer will stand and the applicant will not be considered for subsequent rounds of offers.

10.10 Once an offer has been accepted, the relevant Regional Subcommittee will allocate the successful applicant to a Training Rotation according to the following guidelines:

3.10.1 Allocation will be based on national rank and preference, entry level eligibility (SET 1or SET 2) and the number of available positions.

3.10.2 While every effort will be made to match applicants to their preference, due to the number of posts available this will not always be possible and new trainees are required to accept the rotation allocated to them.

3.10.3 In the interests of fairness, allocations to posts may not be made until after several rounds of offers have been finalised.

10.11 Applicants who have been deemed suitable but who are not ranked highly enough to receive an offer by the final round will be considered unsuccessful. Unsuccessful applicants will be notified in writing as outlined in clause 11 of these Regulations.

4. STRUCTURED CURRICULUM VITAE

4.8. The Structured Curriculum Vitae (online application form) captures information relevant to the eligibility of the applicant and the administration of the selection process, in addition to information on experience, education, research, publications, presentations, development activities and referees.

4.9. Each Structured Curriculum Vitae will be scored by the two (2) people nominated by the Board without reference to the opinions of others using a structured scoring system. For applicants to the Australian program, where any discrepancy occurs in the scores provided by the two (2) scorers, the Board Chairman will score the Structured Curriculum Vitae to identify the anomaly and determine the correct score. For applicants to the New Zealand program, where any discrepancy occurs in the scores the Chair of the New Zealand Subcommittee of the Board will score the Structured Curriculum Vitae to identify the anomaly and determine the correct score.

4.10. The Structured Curriculum Vitae will be scored out of a potential 25 points. The components scored are:

4.3.1 Surgical and Medical Experience (Maximum 5 points)
4.3.2 Skills Courses (Maximum 2 points)
4.3.3 Qualifications (Maximum 6 points)
4.3.4 Presentations (Maximum 4 points)
4.3.5 Publications (Maximum 4 points)
4.3.6 Prizes/Awards for Excellence (Maximum 2 points)
4.3.7 Leadership (Maximum 2 points)

4.11. Surgical and Medical Experience is scored according to the following guidelines, up to a maximum of 5 points:

4.4.1 Scoring will only consider terms undertaken in the last two (2) years except where 4.4.2 applies.

4.4.2 Where the applicant has been undertaking research towards a higher degree in a medically related discipline, scoring will consider terms undertaken in the last four (4) years.

4.4.3 Terms in surgery or a related discipline of less than five (5) weeks will not be scored.

4.4.4 Medical terms of less than ten (10) weeks will not be scored.

4.4.5 Terms planned for after the closing date in the year of application will not be scored.

4.12. Skills Courses are scored according to the following guidelines, up to a maximum of 2 points:

4.5.1 Courses must be delivered by a recognised training provider as determined by the Board.

4.5.2 Scoring includes those related to professional development in clinical and technical competencies including ASSET, CCrISP, EMST, CLEAR and Statistics for Surgeons.

4.5.3 Scoring excludes professional development skills courses that are less than seven (7) hours in duration.

4.5.4 Scoring includes courses related to the development of professional competencies such as communication, teamwork and leadership.

4.13. Qualifications are scored according to the following guidelines, up to a maximum of 6 points:

4.6.1 Scoring only includes higher degrees successfully completed at the time of application at a recognised institution as determined by the Board.

4.6.2 Scoring ·does not include primary medical qualifications including the MBBS/MBChB or overseas equivalent.

4.6.3 Scoring includes Masters degree/s in a medically related area

4.6.4 Scoring includes completion of a PhD.

4.6.5 Scoring includes successful completion of the RACS BSE.

4.14. Presentations are scored according to the following guidelines, up to a maximum of 4 points:

2.3.1 Scoring only includes presentations relevant to medicine.

2.3.2 Scoring only includes presentations personally given by the applicant.

2.3.3 Scoring only includes presentations at scientific meetings or conferences subject to abstract selection.

2.3.4 Presentations that have sufficiently similar topics or that have been presented at more than one scientific meeting or conference will only be scored once.

2.3.5 Presentations will be weighted depending on national, local or international level.

4.15. Publications are scored according to the following guidelines, up to a maximum of 4 points:

i. Scoring only includes publications relevant to medicine.

ii. Scoring only includes publications accepted for publication in a peer reviewed publication and excludes published abstracts.

4.8.3 Each publication can only be scored once.

1.1.1. Scoring includes case reports, articles and book chapters with extra weighting on articles and book chapters where the applicant is the first author.

1.1.2. Documentary evidence of acceptance for publication must be provided at the time of application.

4.16. Prizes/Awards for Excellence are scored according to the following guidelines, up to a maximum of 2 points:

4.9.1 Scoring only includes prizes or awards for excellence in a medically related field, including prizes for presentations.

4.17. Leadership is scored according to the following guidelines, up to a maximum of 2 points:

4.10.1 Applicants may score for an elected or appointed position of responsibility on a board, committee or other appropriate body in a community service or professional organisation, as determined by the Board.

4.10.2 Applicants may score for community and cultural involvement or sporting activities as determined by the Board, with weightings for representation at a state/New Zealand provincial, national or international level.

4.18. The score out of 25 will be adjusted to an overall percentage score rounded to two decimal places for the Structured Curriculum Vitae selection tool.

5. STRUCTURED REFEREE REPORTS

- References are collected to obtain information, in confidence, about the history of the applicant as well as assessments regarding a number of areas such as personal attributes, quality of work and suitability for the SET Program in General Surgery.
- The applicant must provide contact details including a valid email address for the two (2) supervising consultants who had the greatest period of supervision over the applicant for each rotation undertaken in the four (4) clinical years prior to the closing date for applications.

- For applicants undertaking research towards a higher degree in a medically related discipline at the time of application, contact details including a valid email address must be provided for the two (2) supervising consultants who had the greatest period of supervision over the applicant for each rotation undertaken in the four

  (4) years prior to the closing date for applications.
- At a minimum ten (10) referee names must be supplied.
- If an applicant elects not to provide the details for supervising consultants as stipulated by these Regulations, or it is subsequently discovered that the applicant has provided incorrect or misleading information either intentionally or unintentionally, including listing supervising consultants who do not strictly comply with these Regulations, or omitting supervising consultants in preference for others who have had a lesser supervisory role, the applicant may be automatically withdrawn from the selection process and their application will not be considered further in the selection process.
- The units in which the applicant has worked may be contacted as part of the selection process to verify that the supervising consultants listed on the application form comply with these Regulations. The supervising consultants will also be asked to verify compliance with these Regulations.
- The Board will select five (5) supervising consultants to be contacted as part of the selection process. h selecting supervising consultants the Board will endeavour to obtain at least one (1) report from each General Surgery term (where applicable) and the remaining from other terms with consideration given to the duration and type of term.
- The Board will select five (5) alternate supervising consultants to complete the report. Reports completed by alternate supervising consultants will only be used as part of the selection process if one (1) or more of the supervising consultant reports identified in clause 5.7 are not received by the final submission date or if a report is deemed invalid (as in clause 5.15). The alternate supervising consultant reports, where required, will be used in order of their submission date.
- The supervising consultants selected to submit reports will be at the discretion of the Board and the names will not be released to the applicants.
- If, having applied clauses 5.7 and 5.8, the Board has not obtained five (5) valid reports prior to the final submission date determined by the Board, the applicant will be formally

withdrawn from the selection process and their application will not be considered further.

- The Board is responsible for the collection of the reports. Applicants will not be provided with updates on the reports collected; nor will they be involved in the collection process in any way. All supervising consultants contacted as part of the selection process will be advised of the confidential nature of the reports. Harassment of any kind of any individual involved in the completion or collection of the reports is a serious matter and may result in the applicant being deemed unsuitable for selection and removed from the selection process. Harassment includes repeated requests by the applicant to any supervising consultant for a copy of the report submitted.
- On the report the supervising consultant will be asked to select one (1) of five (5) options for each of the twenty (20) assessment areas that they believe best describes the applicant. The selection criteria that will be scored within the reports can be generally categorised as follows:

9.1.1 Medical Expertise (Maximum 8 points)
9.1.2 5.12.2 Technical Expertise (Maximum 24 points)
9.1.3 Judgement/Clinical Decision Making (Maximum 24 points)
9.1.4 Communication {Maximum 16 points)
9.1.5 . Collaboration (Maximum 16 points)
9.1.6 Scholar and Teacher (Maximum 16 points)
9.1.7 Professionalism (Maximum 56 points)

- The options chosen by the referee will be converted to the associated numeric score by the Board using a predetermined scoring system as follows:

6.2.5. The first option is categorised as "unsatisfactory" and scores 0 points.
6.2.6. The second option is categorised as "basic" and scores 2 points.
6.2.7. The third option is categorised as "intermediate" and scores 4 points.
6.2.8. The fourth option is categorised as "advanced" and scores 6 points.
6.2.9. The fifth option is categorised as "expert" and scores 8 points.

- The individual report scores will be converted to a percentage score rounded to two decimal places, calculated by dividing the total score for the report by the total number of questions for which the referee has provided a response.

- If the referee has provided a response for less than 80% of the report, the report will be deemed invalid and will not be used as part of the selection process. h these circumstances an alternate report will be sought (as in clause 5.8).
- The percentage scores for the five (5) individual reports will be combined to provide an overall percentage score, rounded to two decimal places, for the Structured Referee Report selection tool.

6. SEMI-STRUCTURED GENERAL SURGERY PANEL INTERVIEW

2.9 The interview has been designed to:

2.9.1 Identify factors deemed important to the practice of General Surgery.

2.9.2 Address the key competencies as determined by the RACS.

2.9.3 Assess the suitability of the applicant for training.

2.10 The interview seeks information on a variety of attributes including:

2.10.1 The ability to interact effectively and cordially with peers, mentors, members of the health care team, hospital administrators, patients and their families.

2.10.2 The ability to contribute effectively as a member of the health care team.

2.10.3 The ability to act ethically, responsibly and with honesty.

2.10.4 The capacity to care, demonstrate concern and sensitivity to the needs of others.

2.10.5 Effective oral communication.

2.10.6 The ability to assimilate and organise information and to adapt accordingly.

2.10.7 The ability to present concisely within a time frame.

2.10.8 The candidate's commitment to a career in General Surgery.

2.10.9 . The ability to recognise and respond appropriately to ethical issues.

2.10.10 The ability to promote health maintenance and respond to the health needs of the community, patients, colleagues and self.

2.11 Applicants who do not meet the minimum criteria will not be eligible for an interview and will be notified accordingly.

2.12 Applicants will be notified of the date, time and location of the interview at least ten (10) business days prior.

2.13 It is the applicant's responsibility to make the appropriate travel arrangements and to meet any costs incurred in attending the interview. The Board accepts no responsibility for any costs incurred by applicants in attending the interview or applicants who fail to satisfy the minimum standards or eligibility who are not permitted to attend an interview.

2.14 For applicants to the Australian program, interviews will be held in Victoria, Queensland, New South Wales, South Australia and Western Australia. For applicants to the New Zealand program, interviews will be held in Wellington.

2.15 Interview dates will be published on the College website prior to the opening of the application process.

2.16 Applicants must make themselves available at the scheduled interview time. Applicants who do not present for the interview at the scheduled time will not be considered further in the selection process and their application will be withdrawn.

2.17 Candidates will be provided with a brief on the structure of the interview at the time of notification.

2.18 The total score for the Semi-Structured General Surgery Panel Interview selection tool will comprise 40% of the overall selection mark.

The following clauses apply to applicants for the Australian program:

2.19 The interviews will be conducted by a series of five (5) interview panels comprised of two (2) members of the selection committee. Each panel will conduct a designated section of the interview for all applicants, with applicants rotating between panels.

2.20 Candidates will spend 10 minutes with each panel.

2.21 Each interview will be approximately 60 minutes in total duration.

2.22 During the semi-structured interview process, applicants will be asked the same initiating questions by each panel, with follow-up probing questions to explore the breadth and depth of the applicants experience and insight in relation to each selection criterion, particularly as they relate to the nine (9) RACS training competencies.

2.23 Applicants will be scored using a structured scoring system and criterion statements.

2.24 Each panel member will score each applicant individually on a specific form with a consensus score for the interview panel to be arrived at following the interview. The score for each panel will be out of five (5). The consensus score sheet will be used in the final ranking of suitable applicants.

2.25 The total maximum score for answers to the questions will be 25.

2.26 Each panel will also score, out of five (5), the applicant on Communication and Presentation.

2.27 The interview scores from the five (5) panels will then be collated and added to the average for communication score, therefore the interview will be scored out of a potential 30 points.

The following clauses apply to applicants for the New Zealand program:

2.28 The interviews will be conducted by a series of four (4) interview panels comprised of two (2) to three (3) members of the selection committee. Each panel will conduct a designated section of the interview for all applicants, with applicants rotating between panels.

2.29 Candidates will spend approximately 10 minutes with each panel.

2.30 Each interview will be approximately 40 minutes in total duration.

2.31 During the semi-structured interview process, applicants will be asked initiating questions by each panel, with follow-up probing questions to explore the breadth and depth of the applicants experience and insight in relation to each selection criterion, particularly as they relate to the nine (9) RACS training competencies.

2.32 Applicants will be scored using a structured scoring system and criterion statements.

2.33 Each panel member will score each applicant individually on a specific form with a consensus score for the interview panel to be arrived at following the interview. The score for each panel will be out of ten (10). The consensus score will be used in the final ranking of suitable applicants.

2.34 The total maximum score for answers to the questions will be 40.

7. SELECTION COMMITTEE/ PANEL

1.1. The Selection Committee/Panel has been designed to ensure all relevant parties are adequately represented.

1.2. For the Australian Program, each Selection Committee/Panel will comprise two (2) members from the following areas:

2.3.1. Members of the Board in General Surgery.

2.3.2. Members of the Regional Subcommittees of the Board in General Surgery

2.3.3. Fellows of the College who have attended the RACS Interviewer Training Course in the last two (2) years, and who are General Surgeons.
2.3.4. Jurisdictional Representatives.
2.3.5. Hospital Administrators.

1.3. For the New Zealand program, the selection committee will comprise the members of the NZ Regional Subcommittee of the Board in General Surgery (ie. the Hospital supervisors in every training hospital in New Zealand), or their approved alternate.

8. INTERVIEW PANELS

6.2. The interview panels in Australia will be designated one of the following areas:

1.3.1. Scholar and Teacher.
1.3.2. Communication and Collaboration.
1.3.3. Management and Leadership.
1.3.4. Health Advocacy.
1.3.5. Professional and contribution to General Surgery.

6.3. These same areas will be addressed in the New Zealand interviews.

9. INTERVIEW SCORE

6.1. Australian applicants will be scored using the following structured scoring system and criteria:

6.1.1. Unsatisfactory (1 point): The applicant failed to articulate appropriate responses covering some of the key points related to the scoring criteria and did not demonstrate the potential for appropriate knowledge, skills or abilities and/or did not demonstrate some of the personal qualities and behaviours sought.
6.1.2. Basic (2 points): The applicant articulated appropriate responses covering some of the key points related to the scoring criteria and demonstrated the potential for suitable knowledge, skills and abilities with further experience and demonstrated the personal qualities and behaviours sought.
6.1.3. Intermediate {3 points): The applicant articulated appropriate responses covering the key points related to the scoring criteria and demonstrated appropriate knowledge, skills and abilities and the personal qualities and behaviours sought.
6.1.4. Advanced (4 points): The applicant articulated good responses covering all the key points related to the

scoring criteria and demonstrated good knowledge, skills and abilities and the personal qualities and behaviours sought.

6.1.5. Expert (5 points): The applicant articulated excellent responses covering all the key points related to the scoring criteria and demonstrated exceptional knowledge, skills and abilities and the personal qualities and behaviours sought.

6.2. Each question will be accompanied by a criterion answer.

6.3. Interviewers are to score in whole numbers only.

6.4. Australian interviewers are to allocate a score for Communication and Presentation based on the following guidelines:

6.4.1. Unsatisfactory (1 point): The applicant demonstrated no clear organisation in responses provided, core concepts were not integrated into responses, answers did not end in a smooth manner, and speech was unclear and difficult to understand.

6.4.2. Basic (2 points): The applicant's answer was somewhat organised and well thought out, however lost focus regularly, incorporated concepts that were not relevant to the question, conclusion did not flow smoothly within the response, and problems existed with clarity of speech for at least 50% of the presentation.

6.4.3. htermediate (3 points): The applicant presented fairly clearly, however lost focus three to four times, incorporated one or two concepts however these were not relevant to the question, conclusion was well constructed but disjointed from the remainder of the answer, clarity of speech was average but was not confident in answers.

6.4.4. Proficient (4 points): The applicant presented answers that were mostly clear and generally well thought out, however lost focus once or twice, incorporated concepts but missed vital key areas, conclusion was well defined, spoke clearly but demonstrated a lack of confidence once or twice.

6.4.5. Strong (5 points): The applicant presented exceptionally clearly and well thought out responses, remained extremely focussed, incorporated the key concepts, concluded effectively, and spoke clearly and with confidence.

10. FEEDBACK TO UNSUITABLE APPLICANTS

5.12 Applicants who have been deemed unsuitable for selection will not be considered further in the selection process. These applicants will be notified in writing of the following:

5.12.1. That they have been deemed unsuitable for selection and will not be considered further in the selection process.
5.12.2. Information on the overall scores they received for each of the selection tools completed.
5.12.3. Notification of the minimum standard or selection process Regulation that they failed to satisfy.
5.12.4. Should applicants desire further feedback, they may discuss the areas in which they were found to be deficient with their supervisors. Verbal feedback will not be provided.

11. FEEDBACK TO UNSUCCESSFUL APPLICANTS

4.6. Applicants who have been deemed unsuccessful will be notified in writing of the following:

4.6.1. That they have been deemed suitable for selection but have not ranked highly enough to be made an offer in accordance with the intake and have therefore been unsuccessful.
4.6.2. Information on the overall scores received for each of the selection tools completed and their national ranking.
4.6.3. Information on the waiting list process and their position in the list should a position, in accordance with the intake, become available.

12. FEEDBACK TO SUCCESSFUL APPLICANTS

4.7. Applicants who have been deemed successful in the selection process will be notified in writing and by email of the following:

4.7.1. That they have been deemed successful in the selection process and are being offered a position on the SET Program in General Surgery subject to the conditions outlined in clause 12.2.
4.7.2. Information on the overall scores received for each of the selection tools completed and their national ranking.
4.7.3. Information on applicable entry level eligibility (SET1 or SET2) and the Regional allocation and on the process for allocation to a training post.

4.8. Acceptance of the offer to the SET Program in General Surgery will be conditional on the following:

4.8.1. The applicant being registered in the state/territory/country of offer.
4.8.2. The applicant being employed by the relevant health areas and/or the allocated hospital.

4.8.3. The information submitted in the application form being true and correct.

4.8.4. Satisfactory completion of all minimum eligibility criteria by 31 December in the year of application or, for those offered a position on the New Zealand program, before the start of the training year.

4.8.5. Provision of any outstanding documentation required by the Board.

4.8.6. Payment of all monies owed to the College.

4.9. Applicants who fail to satisfy any of the conditions outlined in clause 12.2 of these Regulations will automatically forfeit the offer.

4.10. Applicants who fail to return the acceptance of offer form by the stipulated deadline, or who decline the offer, will automatically forfeit the offer.

4.11. Applicants who return the acceptance of offer form by the stipulated deadline, and who satisfy the conditions outlined in clause 12.2 will be contacted by the relevant Regional Office in accordance with clause 3.10 of these Regulations.

13. DEFERRAL OF TRAINING

4.9. Applicants who wish to defer the commencement of their General Surgical Education and Training must lodge a request to the Board in General Surgery at the time of acceptance using the following procedure:

4.9.1. Lodge the request including any relevant documentation through the "Trainee Online Request" facility on the College website.

4.9.2. The request will be considered by the relevant Regional Subcommittee and forwarded to the Board for approval.

4.10. All applications for deferral or interruption are governed by the SET: Trainee Registration and Variation Policy available on the College website at Policies and Procedures.

4.11. Applicants are required to ensure that their requests for deferral comply with the above policy.

4.12. The Board in General Surgery does not have the authority to alter College Policy, or to approve non-compliant requests.

4.13. Deferrals may not be granted in later rounds of offer due to logistical considerations.

REGULATIONS FOR SELECTION TO SURGICAL EDUCATION AND TRAINING IN GENERAL SURGERY IN NEW ZEALAND

For general instructions and guidelines for selection into the SET Program, please refer to the **SET: Selection to Surgical Education and Training Policy** located on the College Website.

1. INTRODUCTION

a. Definition of terms for the purpose of these Regulations

i. **Applicant** means a person who has submitted an application for the Surgical Education and Training Program in General Surgery to the Royal Australasian College of Surgeons.

ii. **Board** means the Royal Australasian College of Surgeons' Board in General Surgery.

iii. **Business Days** means Monday to Friday excluding Public Holidays.

iv. **College** or **RACS** means the Royal Australasian College of Surgeons.

1.15. **Interview** means the New Zealand Subcommittee of the Board in General Surgery semi-structured General Surgery panel interview conducted as part of the selection process.

1.1.6. **Police Report** means a report on the criminal record of a person.

1.1,7. **Referee** means a person identified in accordance with these Regulations to evaluate professionally the applicant's performance.

1. **Relevant Police Force** means the New Zealand Police Force and / or any or all of the Australian Federal Police and the Australian State and Territory Police forces, as may be applicable.

2. **SET Program** means the Surgical Education and Training Program in General Surgery as approved by the Board in General Surgery.

b. Purpose of these Regulations

The purpose of these Regulations is to set forth and establish the principles, terms and conditions of the selection process for the Royal Australasian College of Surgeons Surgical Education and Training (SET) Program in General Surgery for the 2011 intake. This is a public document.

c. Administration and Ownership

i. The College is the body accredited and authorised to conduct surgical education **and training in Australia and New Zealand and in some regions of Asia.**

ii. The Board in General Surgery is responsible for the delivery of the Surgical Education and Training Program in General Surgery, the accreditation of hospital **posts, and the assessment and supervision of General Surgical Trainees.**

iii. The Board in General Surgery delivers the SET Program in General Surgery in Australia and New Zealand. These regulations apply to New Zealand applicants.

iv. For further information, refer to the **SET Specialty Boards and Regional Subcommittees Terms of Reference** located on the College Website.

2. APPLICATION

d. Applicants wishing to apply to the SET Program in General Surgery in New Zealand must first submit a completed Registration Form to the College via the College website by the published closing date.

e. Applicants are required to confirm for themselves that they meet the minimum eligibility criteria required by the Board before submitting their completed Registration Form. Only applicants who satisfy the eligibility and application requirements in accordance with College policy will be considered in open competition for selection to the SET Program in General Surgery.

f. For further information regarding Registration, including fees, please refer to the **SET: Registration for Selection into the Surgical Education and Training (SET) Policy** available on the College Website.

3.1 Applicants must consent to a full criminal history check including the submission of relevant documentation on request to enable this to be undertaken noting that:

3.1.1 Where consent is not given by the applicant, they will automatically be deemed ineligible for selection and not considered further in the selection process.

3.1.2 Applicants with a relevant criminal conviction will be deemed unsuitable for selection to the training program. A relevant conviction includes, but is not limited to, a conviction of a sexual nature, a conviction relating to drug usage and/or trafficking, a conviction against liberty, morality and abduction, or a conviction relating to dishonesty, fraud and deception.

3.1.3 Failure by an applicant to make full and frank disclosure of their criminal history as requested is grounds to automatically deem the applicant unsuitable for selection, unless the matter is a "spent conviction" under the relevant law.

11.1 Applicants must have current and valid medical registration from the Medical Council of New Zealand at the time of registration. New Zealand applicants must have general scope registration without restriction or general scope registration restricted to general surgery.

11.2 Applicants must have citizenship or have been granted permanent residency status in New Zealand or Australia at the time of registration.

11.3 Applications can only be submitted via the College online application system at www.surqeons.org. No other form of application will be accepted and no extensions will be granted. It is the applicant's responsibility to ensure that they allow enough time to complete the application.

11.4 Separate applications must be made for the SET Program in General Surgery in Australia and the SET Program in General Surgery in New Zealand. Applicants can not apply for both programs.

11.5 Applicants must note the following General Surgery specific eligibility requirement:

Rotation Type	Minimum Duration	Validity Period	Completed By
General surgery rotation	1x 8 week	2 years, extended up to 4 years by a period of full time study in a medically related discipline.	By the end of 2010
Surgery in general (Surgery in any surgical discipline / sub-specialty),	1x 8 weeks	2 years, extended up to 4 years by a period of full time study in a medically related discipline.	By the end of 2010
Surgery in critical care. (refer to 2.9.1 for Definition of a Critical Care Term)	1X 8 week		By the end of 2010

11.5.1 Examples of a Critical Care term are as follows

11.5.1.1 Trauma Unit
11.5.1.2 ICU
11.5.1.3 HDU
11.5.1.4 ED
11.5.1.5 Cardiothoracic Unit
11.5.1.6 Vascular Unit
11.5.1.7 Burns Unit
11.5.1.8 Anaesthetic Unit
11.5.1.9 Transplant / HPB

11.5.1.10 Critical Care Unit

11.5.2 Surgical Terms cannot be considered for more than one eligibility requirement. Applicants will need to stipulate if the term is to be considered as general surgery, surgery in general or critical care.

11.5.3 Applicants must provide proof of past and future rotations in the form of a letter of confirmation from the hospital or copy of the applicable roster.

3. SELECTION PROCESS OVERVIEW

1. Applicants who satisfy the eligibility and application requirements in accordance with College policy will be considered in open competition for selection to the SET Program in General Surgery.
2. On completion of the relevant components of the selection process, eligible applicants will be classified as one of the following:
    1. **Unsuitable** being an eligible applicant who failed to satisfy a minimum standard for selection.
    2. **Unsuccessful** being an eligible applicant who satisfied the minimum standards for selection who is therefore suitable but who did not rank highly enough in comparison to the intake to be made an offer of selection.
    3. **Successful** being an eligible applicant who satisfied the minimum **standards** for selection who is therefore suitable and who has ranked highly enough in comparison to the intake to be made an offer of selection.
3. Applicants who satisfy the minimum standard for selection will be deemed suitable for selection and will be ranked. The ranking will be determined by applying the following weightings to the percentage adjusted score out of 100 obtained for each of the three (3) selection tools, providing an overall percentage score:

7.1.1 Structured Curriculum Vitae 20%

7.1.2 Structured Referee Reports 40%

7.1.3 Semi-Structured General Surgery Panel Interviews 40%

4. Applicants will be offered positions on their national ranking only.
5. It is expected that due to attrition and requests for interruption / deferral, there will be several rounds of offers to the SET Program in General Surgery.
6. Applicants who have been deemed suitable but who are not ranked highly enough to receive a first round offer to the SET Program in General Surgery, will still be considered eligible for subsequent rounds of offers made by the New Zealand Subcommittee of the Board.

7. Applicants who do not wish to receive a later round offer to the SET Program in General Surgery must advise the New Zealand Subcommittee of the Board by the stipulated deadline.

8. Once an offer has been accepted, the New Zealand Subcommittee of the Board will allocate the successful applicant to a Training Rotation according to the following guidelines:

   1. Allocation will be based on preference, entry level eligibility (SET 1or SET 2) and the number of available positions.

   2. While every effort will be made to match applicants to their preference, due to the number of posts available this will not always be possible and new trainees are required to accept the rotation allocated to them.

   3. In the interests of fairness, allocations to posts may not be made until several rounds of offers have been finalised.

9. Applicants who have been deemed suitable but who are not ranked highly enough to receive an offer by the final round will be considered unsuccessful. Unsuccessful applicants will be notified in writing as outlined in clause 11 of these Regulations.

4. STRUCTURED CURRICULUM VITAE

6.1 The Structured Curriculum Vitae (online application form) captures information relevant to the eligibility of the applicant and the administration of the selection process, in addition to information on experience, education, research, publications, presentations, development activities and referees.

6.2 Each Structured Curriculum Vitae will be scored by two (2) members of the New Zealand Subcommittee of the Board without reference to the opinions of others using a structured scoring system. Where any discrepancy occurs in the scores provided by the two (2) scorers, the Chair of the New Zealand Subcommittee of the Board (or his / her delegate) will score the Structured Curriculum Vitae to identify the anomaly and determine the correct score.

6.3 The Structured Curriculum Vitae will be scored out of a potential 25 points. The components scored are:

5.1.1 Surgical and Medical Experience (Maximum 7 points)

5.1.2 Skills Courses (Maximum 2 points)

5.1.3 Qualifications (Maximum 4 points)

5.1.4 Presentations and Publications (Maximum 5 points)

5.1.5 Prizes/Awards for Excellence (Maximum 2 points)

5.1.6 Leadership (Maximum 2 points)

5.1.7 Scholar and Teacher (Maximum 3 points)

6.4 Surgical and Medical Experience is scored according to the following guidelines, up to a maximum of 7 points:

12.1.1 Scoring will only consider terms undertaken in the last two (2) years except where 4.4.2 applies.

12.1.2 Where the applicant has been undertaking full time research towards a higher degree in a medically related discipline, scoring will consider terms undertaken in the last four (4) years.

12.1.3 Terms in surgery or a related discipline of less than five (5) weeks will not be scored.

12.1.4 4.4.4. Medical terms not of a surgical nature will not be scored.

12.1.5 4.4.5. Terms planned for after the closing date in the year of application will not be scored.

6.5 Skills Courses are scored according to the following guidelines, up to a maximum of 2 points:

12.3.1 Scoring will consider courses undertaken in the past five (5) years.

12.3.2 Courses must be completed at the time of application closing date and must be accompanied by documentation as evidence of attendance / completion.

12.3.3 Courses must be delivered by a recognised training provider as determined by the Board.

12.3.4 Scoring includes those related to professional development in clinical and technical competencies including ASSET, CCr!SP, EMST, CLEAR and Statistics for Surgeons.

12.3.5 Scoring excludes professional development skills courses that are less than seven (7) hours in duration.

12.3.6 Scoring includes courses related to the development of professional competencies such as communication, teamwork and leadership.

6.6 Qualifications are scored according to the following guidelines, up to a maximum of 4 points:

4.10.1 Scoring only includes higher degrees successfully completed at the time of
application at a recognised institution as determined by the Board.

4.10.2 Scoring does not include primary medical qualifications including the MBChB / MBBS or overseas equivalent.

4.10.3 Scoring includes Masters degree/s in a medically related area.

4.10.4 Scoring includes completion of a PhD.

4.10.5 Scoring includes successful completion of the RACS Basic Surgical Examination.

4.10.6 Scoring does not include the MRCS.

4.10.7 Documentary evidence of completion must be provided at the time of application

6.7 Presentations and Publications are scored according to the following guidelines, up to a maximum of 5 points:

1.1.1. Scoring will consider presentations or publications undertaken in the past five (5) years**.**

1.1.2. Presentations and publications must be complete at the time of application closing date. Prospective presentations and publications will not be scored.

1.1.3. Scoring only includes presentations relevant to surgery.

1.1.4. Scoring only includes presentations personally given by the applicant.

1.1.5. Scoring only includes presentations at scientific meetings or conferences subject to abstract selection.

1.1.6. Presentations that have sufficiently similar topics or that have been presented at more than one scientific meeting or conference will only be scored once.

1.1.7. Presentations will be weighted depending on national, local or international level.

1.1.8. Scoring only includes publications relevant to surgery.

1.1.9. Scoring only includes publications accepted for publication in a peer reviewed publication and excludes published abstracts.

1.1.10. Each publication can only be scored once.

1.1.11. Scoring includes case reports, articles and book chapters with extra weighting on articles and book chapters where the applicant is the first author.

1.1.12. Documentary evidence of acceptance for publication and proof of presentation must be provided at the time of application.

6.8 Prizes/Awards for Excellence are scored according to the following guidelines, up to a maximum of 2 points:

4.8.1 Scoring only includes prizes or awards for excellence in a medically related field, including prizes for presentations.

4.8.2 Documentary evidence of award or prize must be provided at the time of application.

6.9 Leadership is scored according to the following guidelines, up to a maximum of 2 points:

1.1.8. Applicants may score for an elected or appointed position of responsibility on a board, committee or other appropriate body in a community service or professional organisation, as determined by the Board.

1.1.9. Applicants may score for community and cultural involvement or sporting activities as determined by the Board.

1.1.10. Evidence of involvement from the relevant institution must be supplied.

6.10 Scholar/Teacher is scored according to the following guidelines, up to a maximum of 3 points:

- Applicants may score for involvement in continued teaching.
- Evidence of involvement from the relevant institution must be supplied.

6.11 The score out of 25 will be adjusted to an overall percentage score out of 20 rounded to two decimal places for the Structured Curriculum Vitae selection tool.

5. STRUCTURED REFEREE REPORTS

a. References are collected to obtain information, in confidence, about the history of the applicant as well as assessments regarding a number of areas such as personal attributes, quality of work and suitability for the SET Program in General Surgery.

b. The applicant must provide the names of the supervising consultants (up to a maximum of three (3) consultants per rotation) who had the greatest period of supervision over the applicant for each rotation undertaken in the two (2) clinical years prior to the closing date for applications. Applicants who have been undertaking a period of full time study in a medically related discipline within the previous two years may extend that period by the period of the full time study, up to a maximum of four (4) years.

c. Applicants must select ten (10) referee names from the supervising consultants named.

d. If an applicant elects not to provide the details for supervising consultants as stipulated by these Regulations, or it is subsequently discovered that the applicant has provided incorrect or misleading information either intentionally or unintentionally, including listing supervising consultants who do not strictly comply with these Regulations, or omitting supervising consultants in preference for others who have had a lesser supervisory role, the applicant may be automatically withdrawn from the selection process and their application will not be considered further in the selection process.

e. The units in which the applicant has worked may be contacted as part of the selection process to verify that the supervising consultants listed on the application form comply with these Regulations. The supervising consultants will also be asked to verify compliance with these Regulations.

f. The New Zealand Subcommittee of the Board will select five (5) supervising consultants from the ten (10) referees named to be contacted as part of the selection process. In selecting supervising consultants the New Zealand Subcommittee will endeavour to obtain at least one (1) report from each General Surgery term (where applicable) and the remaining from other terms with consideration given to the duration and type of term.

g. The remaining five (5) will be alternate supervising consultants to complete the report. Reports completed by alternate supervising consultants will only be used as part of the selection process if one (1) or more of the supervising consultant reports identified in clause 5.6 are not received by the final submission date or if a report is deemed invalid (as in clause 5.14). The alternate supervising consultant reports, where required, will be used in order of their submission date.

h. The supervising consultants selected to submit reports will be at the discretion of the Board and the names will not be released to the applicants.

i. If, having applied clauses 5.6 and 5.7, the Board has not obtained five (5) valid reports prior to the final submission date determined by the New Zealand Subcommittee of the Board, the applicant will be formally withdrawn from the selection process and their application will not be considered further.

j. The New Zealand Subcommittee of the Board is responsible for the collection of the reports. Applicants will not be provided with updates on the reports collected; nor will they be involved in the collection process in any way. All supervising consultants contacted as part of the selection process will be advised of the confidential nature of the reports. Harassment of any kind of any individual involved in the completion or collection of the reports is a serious matter and may result in the applicant being deemed unsuitable for selection and removed from the selection process. Harassment includes repeated requests by the applicant to any supervising consultant for a copy of the report submitted.

k. On the report the supervising consultant will be asked to select one (1) of four (4) options for each of the sixteen (16) assessment areas that they believe best describes the applicant. The selection criteria that will be scored within the reports can be generally categorised as follows:

i. Medical and Technical Expertise

ii. Judgement/Clinical Decision Making

iii. 5.11.3. Communication

iv. Collaboration

v. Scholar and Teacher

vi. Professionalism

l. The options chosen by the referee will be converted to the associated numeric score by the New Zealand Subcommittee of the Board using a predetermined scoring system as follows:

2.4.1. The first option is categorised as "unsatisfactory" and scores 0 points.

2.4.2. The second option is categorised as "basic" and scores 2 points.

2.4.3. The third option is categorised as "intermediate" and scores 4 points.

2.4.4. The fourth option is categorised as "advanced" and scores 6 points.

m. The individual report scores will be converted to a percentage score rounded to two decimal places, calculated by dividing the total score for the report by the total number of questions for which the referee has provided a response.

n. If the referee has provided a response for less than 80% of the report, the report will be deemed invalid and will not be used as part of the selection process. In these circumstances an alternate report will be sought (as in clause 5.7).

o. The percentage scores for the five (5) individual reports will be combined to provide an overall percentage score, rounded to two decimal places, for the Structured Referee Report selection tool.

p. The referee reports will comprise 40% of the total selection score.

6. SEMI-STRUCTURED GENERAL SURGERY PANEL INTERVIEW

a. The interview has been designed to:

2.5.1. Identify factors deemed important to the practice of General Surgery.

2.5.2. Address the key competencies as determined by the RACS.

2.5.3. Assess the suitability of the applicant for training.

b. The interview seeks information on a variety of attributes including:

3.3.1. The ability to interact effectively and cordially with peers, mentors, members of the health care team, hospital administrators, patients and their families.

3.3.2. The ability to contribute effectively as a member of the health care team.

3.3.3. The ability to act ethically, responsibly and with honesty.

3.3.4. The capacity to care, demonstrate concern and sensitivity to the needs of others.

- Effective oral communication.
- The ability to assimilate and organise information and to adapt accordingly.
- The ability to present concisely within a time frame.
- The candidate's commitment to a career in General Surgery.
- The ability to recognise and respond appropriately to ethical issues.

- The ability to promote health maintenance and respond to the health needs of the community, patients, colleagues and self.

c. Applicants will be notified of the date, time and location of the interview at least ten (10) business days prior.

d. It is the applicant's responsibility to make the appropriate travel arrangements and to meet any costs incurred in attending the interview. The New Zealand Subcommittee of the Board accepts no responsibility for any costs incurred by applicants in attending the interview.

e. Interviews will be held in Wellington.

f. Interview date(s) will be published on the College website prior to the opening of the application process.

g. Applicants must make themselves available at the scheduled interview time. Applicants who do not present for the interview at the scheduled time will not be considered further in the selection process and their application will be withdrawn.

h. Candidates will be provided with a brief on the structure of the interview at the time of notification.

i. The total score for the Semi-Structured General Surgery Panel Interview selection tool will comprise 40°/o of the overall selection mark.

j. The interview will be conducted by a series of five (5) interview panels comprised of two (2) to three (3) members of the selection committee. Each panel will conduct a designated section of the interview for all applicants, with applicants rotating between panels.

k. Candidates will spend approximately 10 minutes with each panel.

l. The semi structured interview will be approximately 50 minutes in total duration.

m. During the semi-structured interview process, applicants will be asked initiating questions by each panel, with follow-up probing questions to explore the breadth and depth of the applicants experience and insight in relation to each selection criterion, particularly as they relate to the nine (9) RACS training competencies.

n. Applicants will be scored using a structured scoring system and criterion statements.

o. Each panel member will score each applicant individually on a specific form with a consensus score for the interview panel to be arrived at following the interview, The score for each panel will be out of ten (10). The consensus score will be used in the final ranking of suitable applicants.

p. The total score out of 50 will be adjusted to an overall percentage score out of 40 rounded to two decimal places for the Semi Structured Interview selection tool.

7. SELECTION COMMITTEE/ PANEL

1. The Selection Committee/Panel has been designed to ensure all relevant parties are adequately represented.
2. For the New Zealand program, the selection committee will comprise the members of the New Zealand Subcommittee of the Board in General Surgery (ie. the Hospital Supervisors in every training hospital in New Zealand), or their approved alternate.

8. INTERVIEW PANELS

1. These areas will be addressed in the New Zealand interviews:
    1. Scholar and Teacher.
    2. Communication and Collaboration.
    3. Management and Leadership.
    4. Professionalism and contribution to general surgery
    5. Health advocacy and cultural awareness

9. INTERVIEW SCORE

4.3. New Zealand applicants will be scored using the following structured scoring system and criteria:

4.3.1. Unsatisfactory (2 point): The applicant failed to articulate appropriate responses covering some of the key points related to the scoring criteria and did not demonstrate the potential for appropriate knowledge, skills or abilities **and/ or** did not demonstrate some of the personal qualities and behaviours sought.

4.3.2. Basic (4 points): The applicant articulated appropriate responses covering some of the key points related to the scoring criteria and demonstrated the potential for suitable knowledge, skills and abilities with further experience **and** demonstrated the personal qualities and behaviours sought.

4.3.3. htermediate (6 points): The applicant articulated appropriate responses covering the key points related to the scoring criteria and demonstrated appropriate knowledge, skills and abilities **and** the personal qualities and behaviours sought.

4.3.4. Advanced (8 points): The applicant articulated good responses covering all the key points related to the scoring criteria and demonstrated good knowledge, skills and abilities **and** the personal qualities and behaviours sought.

4.3.5. Expert (10 points): The applicant articulated excellent responses covering all the key points related to the scoring criteria and demonstrated exceptional knowledge, skills and abilities and the personal qualities and behaviours sought.

4.4. Each question will be accompanied by a criterion answer.

4.5. Interviewers are to score in whole numbers only.

4.6. In awarding points, interviewers will also consider the applicant's ability to present clear and well thought out responses, remain focused on the question(s), incorporate key concepts and respond in an understandable manner.

10. FEEDBACK TO UNSUITABLE APPLICANTS

4.4. Applicants who have been deemed unsuitable for selection will not be considered further in the selection process. These applicants will be notified in writing of the following:

4.4.1. That they have been deemed unsuitable for selection and will not be considered further in the selection process.

4.4.2. Information on the overall scores they received for each of the selection tools completed.

4.4.3. Notification of the minimum standard or selection process Regulation that they failed to satisfy.

4.4.4. Should applicants desire further feedback, they may discuss the areas in which they were found to be deficient with their supervisors.

11. FEEDBACK TO UNSUCCESSFUL APPLICANTS

1. Applicants who have been deemed unsuccessful will be notified in writing of the following:

    1. That they have been deemed suitable for selection but have not ranked highly enough to be made an offer in accordance with the intake and have therefore been unsuccessful,

    2. Information on the overall scores they received for each of the selection tools completed.

    3. Information on the waiting list process.

2. Applicants who have been deemed suitable but who are not ranked highly enough to receive an offer by the final round will be considered unsuccessful. Such applicants will be notified in writing and should they desire further feedback, they may discuss the

information on their overall scores for each of the selection tools with their supervisors.

12. FEEDBACK TO SUCCESSFUL APPLICANTS

a. Applicants who have been deemed successful in the selection process will be notified in writing and by email of the following:

1. That they have been deemed successful in the selection process and are being offered a position on the SET Program in General Surgery subject to the conditions outlined in clause 12.2.
2. Information on applicable entry level eligibility (SET1 or SET2) and on the process for allocation to a training post.
3. The due date by which their Offer Form must be returned.

b. The Offer Form has three (3) options - accept, decline or pending

1. Accept - the applicant accepts the offer of a position on the General Surgery program
2. Decline - the applicant declines the offer of a position on the General Surgery program
3. Pending - the applicant wishes to await the outcome of an application to a bi- national surgical training program before deciding on the General Surgery offer.
4. An applicant who selects "Pending" must advise the NZ Subcommittee of the Board whether s/he accepts or declines the General Surgery offer by the date stipulated on the Offer Form. This date is approximately one week after the offers are released for the bi-national programs.

c. Acceptance of the offer to the SET Program in General Surgery will be conditional on the following:

i. The applicant having the appropriate medical registration in New Zealand.

ii. The applicant being employed by the relevant District Health Board.

iii. The information submitted in the application form being true and correct.

iv. Satisfactory completion of all minimum eligibility criteria before the start of the training year in New Zealand.

v. Provision of any outstanding documentation required by the Board.

vi. Provision of a signed "Training Agreement"

vii. Payment of all monies owed to the College.

d. Applicants who fail to satisfy any of the conditions outlined in clause 12.3 of these Regulations will automatically forfeit the offer.

e. Applicants who fail to return the acceptance of *offer* form by the stipulated deadline, or who decline the offer, will automatically forfeit the offer.

f. Applicants who accept a position on the General Surgery program will be allocated to a training post in accordance with clause 3.8 of these Regulations.

13. DEFERRAL OF TRAINING

13.1 Applicants who wish to defer the commencement of their General Surgical Education and Training must lodge a request to the New Zealand Subcommittee of the Board at the time of acceptance using the following procedure:

4.5.1. Complete the required section on the Offer Form

4.5.2. The request will be considered by the New Zealand Subcommittee and forwarded to the Board for final decision.

13.2 All applications for deferral or interruption are governed by the SET: Trainee Registration and Variation Policy available on the College website.

13.3 Applicants are required to ensure that their requests for deferral comply with the above policy.

13.4 The Board in General Surgery does not have the authority to alter College Policy, or to approve non-compliant requests.

13.5 Deferrals will not be granted within three (3) months of the start of the training year due to logistical considerations.

APPENDIX D – RACS and General Surgery selection instruments

One General Surgery CV and one Referee Report are presented sa examples:

1. General Surgery CV 2008

2. General Surgery Referee Report 2010

Applicant: «PREFIX» «FIRST_NAME» «LAST_NAME» («ID»)

Medical Expertise and Technical Expertise				
Surgical and Medical Experiences				
1. Private assisting terms are not scored 2. Terms which are not undertaken on a full time basis will be adjusted pro rata 3. Scoring will only consider terms undertaken in the last 2 years or the last 4 years where an applicant is undertaking research towards a higher degree in a medically related discipline at the time of application. 4. Terms planned for after the closing date in the year of application are not scored 5. A term in medicine is scored 1 point per 10 weeks 6. A term of surgery in general is scored 1 point per 5 weeks				
Maximum 8 points	Scorer 1	Scorer 2	Chair (if required)	**Comments:** «CDC»
Skills Courses				
1. Courses must be delivered by a recognised training provider as determined by the Board 2. Scoring does not include hospital based courses and meetings or activities less than five hours in duration 3. Scoring for courses includes those related to professional development in clinical and technical competencies 4. Scoring for courses includes those related to the development of professional competencies 5. Morbidity and mortality meetings are not scored 6. Each course is scored 1 point				
Maximum 2 points	Scorer 1	Scorer 2	Chair (if required)	**Comments:** «CDC»
Scholar and Teacher				
Qualifications				
1. Scoring only includes recognised higher degrees successfully completed at the time of application 2. Scoring does not include primary medical qualifications (MBBS or overseas equivalent) 3. A Masters degree is scored 1 point for coursework or 2 points for Masters by thesis				
Maximum 3 points	Scorer 1	Scorer 2	Chair (if required)	**Comments:** «CDC»
Presentations				
1. Scoring only includes presentations and posters relevant to medicine 2. Presentations which have sufficiently similar topics or that have been presented at more than one scientific meeting or conference will only be scored once				
Maximum 4 points	Scorer 1	Scorer 2	Chair (if required)	**Comments** «PREC»
Publications				

1. Scoring only includes publications relevant to medicine
2. Scoring only includes publications accepted for publication in a peer reviewed publication and excludes abstracts
3. Each publication can only be scored once
4. A case report is scored 1 point
5. A peer reviewed journal article or book chapter where the applicant is not the first author is scored 2 points
6. A peer reviewed journal article or book chapter where the applicant is the first author is scored 3 points

Maximum 4 points	Scorer 1	Scorer 2	Chair (if required)	**Comments:** «PC»

Prizes and Awards

1. Applicants may score 1 point for prize awarded for a presentation relevant to surgery.
2. Applicants may score 1 point for a prize received for academic achievement in surgery at an undergraduate or post-graduate level.

Maximum 2 points	Scorer 1	Scorer 2	Chairman (if required)	**Comments:** «PC»

Management & Leadership

Leadership and Community Service

1. Applicants may score 1 point for an elected or appointed position of responsibility on a board, committee or other appropriate body as determined by the Board in a community service or professional organisation.
2. Applicants may score 1 point for community involvement by undertaking volunteer work with community groups or community projects on a regular basis.

Maximum 2 points	Scorer 1	Scorer 2	Chairman (if required)	**Comments:** **«PC»**

Overall Score

Maximum 25 points	Scorer 1	Scorer 2	Chairman (if required)	**Do not complete: Admin Use Only** Percentage Score: Minimum Standard Achieved ☐Yes ☐No

Certification

Scorer One Signature: ________________ **Scorer Two Signature:** ________________

Scorer One Name: ____________ **Scorer Two Name:** ________________

Scorer One Date: ____ / ____ / 2008 **Scorer One Date:** ____ / ____ / 2008

Selection Structured Referee Report Surgical Education and Training 2010

Applicant Name: __________________Referee Name: _________________________

How were you selected to provide a referee report?
Applicants for the Surgical Education and Training Programs are required to provide the names and contact details of potential referees. For some specialties the applicants were required to provide the names of the supervising consultants who had the greatest period of supervision over the applicant in each clinical rotation. In these instances the selection or exclusion of supervising consultants was not at the discretion of the applicant and the specialty selected referees independently from those listed. For some specialties the applicants were required to nominate referees or other individuals who were able to provide an assessment of their performance. Your name has been provided in one or more of these capacities.

What SET Programs will this report apply to?
The following specialties have agreed to use a generic structured referee report. Where an applicant has applied to one or more of these specialties and the specialties have selected you as a referee your report will be used. This avoids referees having to complete multiple reports for the same applicant for the specialties below.

Cardiothoracic Surgery
General Surgery
Neurosurgery
Otolaryngology – Head and Neck Surgery
Plastic Surgery
Paediatric Surgery
Urology
Vascular Surgery

Is the assessment provided confidential?
Assessment reports received as part of the SET Program selection process are collected **in confidence**. Applicants are not involved in the collection process. Reports are confidential in nature and should not be released to applicants. Harassment by any applicant, including repeated requests by the applicant for a copy of the report submitted is a serious matter and may result in the applicant's immediate removal from the selection process.

What is the role of the report?
The reports obtain information, **in confidence**, about the history of the applicant's performance, as well as assessments regarding a number of areas such personal attributes, quality of work and suitability for surgical training. The College would appreciate your careful consideration of the selections made as they will influence the overall ranking of the applicants in what is a highly competitive process **so it is important to give a fair and accurate account of performance.**

How do I complete the report?

<u>In all instances, the answer should relate to what is expected of doctors of a similar seniority and experience</u>.

For each attribute the four behavioral statements indicate a level of proficiency that ranges from unsafe, through varying degrees of skill that point to the applicant's readiness to begin training as a surgeon. Please select the option for each attribute which corresponds to the statement you believe **best describes** the applicant's demonstrated skill or behaviour in relation to experience. It is extremely rare for anyone to perform consistently across the behaviours in every attribute – it is also extremely unlikely that they will perform at the highest level in more than a small number of attributes. Therefore **please consider each attribute in isolation, thinking of examples of behaviour to support your response.**

<u>If you have not observed the behaviours relating to one of the attributes please leave that attribute and move on to the next one.</u>

Questions?
Please contact us at any time if you have concerns or questions on + 61 3 9249 1114 or via email at SETenquiries@surgeons.org.

DECLARATION OF CONFLICT OF INTEREST

Note: A declaration of Conflict of Interest will be taken into consideration but will not necessarily make a Referee Report invalid. However, if this section is not completed, the Referee Report will be declared invalid

Do you have any conflict of interest? ☐Yes
☐No

(An apparent conflict of interest is one in which a reasonable person would think that the professional's judgment is likely to be compromised. This may be either a personal or family relationship, or involvement in actions of a disciplinarily nature)
If yes: please provide details

For each assessment area please tick <u>one box</u> only corresponding to the statement you believe <u>best describes</u> the applicant's behavior. ***If you have not observed the behaviours relating to one of the attributes please leave that attribute and move on to the next one.***

MEDICAL EXPERTISE & TECHNICAL EXPERTISE

1 Clinical expertise

☐ Poor knowledge of basic sciences; Difficulty in applying apply basic sciences to clinical situations; Fails to seek or explore relevant information / signals; Suggests too narrow a range of options; Overlooks important issues; Fails to identify risks

☐ Knowledge of basic sciences adequate; Generally able to apply basic sciences to clinical situations

☐ Knowledge of basic sciences adequate; Able to apply basic sciences to clinical situations; Identifies and elicits necessary information; Identifies appropriate options

☐ Comprehensive knowledge of basic sciences; Consistently effectively applies basic sciences to clinical situations; Identifies and elicits necessary information; Identifies appropriate options; Identifies key issues involved well – gets to the root cause; Identifies potential risks

2 Technical ability **Please note:** *Your assessment on attributes should be in relation to the stage and level of experience ie a PGY2 would not be expected to have high levels of technical expertise.*

☐ Sometimes handles tissues and instruments in a risky and/or unsafe manner; Lacks awareness of their own technical limitations; Sometimes lacks manual dexterity or hand-eye coordination; Slow to learn technical /procedural skills;

- [ ] Technical skills performed too hasty, too slowly or without efficiency; Unable to anticipate possible issues
- [ ] Handles tissues and instruments to minimize risk and maximize safety; Generally aware of own technical limitations
- [ ] Handles tissues and instruments to minimize risk and maximize safety; Aware of own technical limitations; Manual dexterity and hand-eye coordination appropriate to perform required tasks; Keen and quick to learn
- [ ] Manual dexterity and hand-eye coordination appropriate to perform required tasks; Performs technical skills in an orderly, efficient and systematic manner; Handles tissues and instruments to minimize risk and maximize safety; Aware of own technical limitations; Keen and quick to learn; Anticipates possible issues

JUDGEMENT – CLINICAL DECISION MAKING

3 Judgement under Pressure

- [ ] Poor decision making under pressure; Tries to deal with a situation alone – to the risk of the patient; Becomes tense or agitate under pressure; Tends to be disorganised; Incapable of making decisions, or tends to make poor decisions; Uneasy with patier emotions / questions;
- [ ] Generally makes correct decisions under pressure; Is aware of own limitations and tends to seeks help appropriately;
- [ ] Makes correct decisions under pressure with ease; Is aware of own limitations and tends to seeks help appropriately; Uses effective strategies to deal with pressure; Remains calm – relaxed and comfortable with the demands of the situation;
- [ ] Clear thinking, rapidly comes to correct clinical decision under pressure; Recognises own limitations and seeks help when necessary; Uses effective strategies to deal with pressure; Remains calm – relaxed and comfortable with the demands of the situation; Well organised and clear planning evident; Uses effective strategies to deal with patient's emotions and questions;

4 Situation awareness

- [ ] Lacks attention to and/or awareness of potential changes; Unable /struggles to adapt their thinking to changing demands of situation; Lacks awareness of level of competence of other team members and/or their capacity to contribute in a clinical situation; Tends to focus on minutia and loose perspective; Unable to anticipate potential problems; Fails to pick up on signs indicate possible change;
- [ ] Alert to symptoms and signs suggesting conditions that might change; Generally responds flexibly – able to adapt thinking to changing situation demands
- [ ] Alert to symptoms and signs suggesting conditions that might change; Able to adapt thinking to changing situation demands; Aware of team members and their capacity to contribute in a clinical situation; Able to maintain wider perspective at the sar time as attending to details;
- [ ] Alert to symptoms and signs suggesting conditions that might change; Responds flexibly – able to adapt thinking to changing situation demands; Aware of team members and their capacity to contribute in a clinical situation; Able to maintain wider perspective at the same time as attending to details; Thinks ahead and anticipates potential changes; Shows vigilance and awareness of subtle changes;

5 Problem solving

- [ ] Unable to identify an appropriate solution; Focuses on peripheral issues; Makes immediate assumptions; Lacks logic in their reasoning; Deals with issues narrowly/ dogmatically; Tends to overlook complexity or ambiguity;
- [ ] Able to identify an appropriate solution; Generally identifies and focuses on key points
- [ ] Able to identify an appropriate solution; Identifies and focuses on key points; Ensures that all relevant information is considered; Clear and rational approach to difficult issues
- [ ] Generates appropriate and practical solutions; Identifies and focuses on key points; Ensures that all relevant information is considered; Clear and rational approach to difficult issues; Seeks best approach for each problem; Anticipates ambiguity and seeks evidence;

6 Decision making

- [ ] Fails to consider all of the facts; Unwilling to take decision and/ or makes inappropriate decisions; Does not consider, or know the merit of, different options; Fails to seek additional information and /or advice; Little understanding of the complexity of the issue/task; Fails to recognise the significance of findings;
- [ ] Generally considers relevant facts; Makes timely and appropriate decisions;
- [ ] Considers all of the facts; Makes timely and appropriate decisions; Identifies and knows the merit of different options; Seeks additional information and / or advice appropriately;
- [ ] Considers all of the facts; Makes timely and appropriate decisions; Identifies and knows the merit of different options; Seeks additional information and /or advice appropriately; Recognises and adapts decision making to match the complexity of the issue / task; Consistently identifies the significance of findings;

7 Organisation and Planning

- [ ] Fails to meet reasonable deadlines; Struggles to prioritise conflicting demands; Has problems in thinking ahead, pre-planning and / or building effective contingencies; Organisation / planning is difficult for others to follow; Unsystematic approach to dealing with time and/ or information; Does not use available resources effectively;
- [ ] Meets reasonable deadlines; Generally prioritises conflicting demands
- [ ] Meets reasonable deadlines; Prioritises conflicting demands well; Thinks ahead and plans effectively for contingencies or possible changing demands; Organisation / planning is easy for others to follow;

☐ Meets reasonable deadlines; Effectively prioritises conflicting demands; Thinks ahead and plans effectively for contingencies or possible changing demands; Organisation / planning is easy for others to follow; Systematic approach to dealing with organisational and planning issues; Makes efficient use of available resources;

COMMUNICATION

8 Communication with colleagues and team members

☐ Fails to keep team members up to date in a timely manner; Poor / inadequate written communication; Poor / inadequate verbal communication; Has poor relationships with peers; Fails to provide clear directions and descriptions of situations to team members; Is defensive or uncompromising when questioned by other staff;

☐ Generally keeps all team members up to date without prompting; Satisfactory written communication

☐ Keeps all team members up to date without prompting; Effective and timely written communication; Effective verbal communication; Has good relationships with peers;

☐ Effective and timely written communication; Effective verbal communication; Keeps all team members up to date without prompting; Has excellent relationship with peers; Always provides clear directions and descriptions of situations; Remains flexible and open when questioned by other staff;

9 Communication with patients

☐ Avoids communication with patients and families where possible; Does not adapt communication to suit the situation; Little use of active listening skills; Over-use of closed questions; Shows little evidence of understanding patient's questions or feelings; Non-verbal behaviour does not facilitate communication;

☐ Communicates with patients and families where possible; Adapts communication as appropriate

☐ Communicates with patients and families where possible; Adapts communication as appropriate; Consistently uses active listening skills; Effective use of open questions;

☐ Develops rapport and effective communication with patients and their families; Consistently uses active listening skills; Effective use of open questions; Adapts communication as appropriate; Responds to patient questions and concerns appropriately; Makes effective use of non-verbal behaviour;

COLLABORATION

10 Team involvement

☐ Had strained relationships with other team members; Does not respond positively to direction or supervision; Displays uncooperative or negative behaviour towards others; Unwilling to negotiate or compromise appropriately; Tends to be critical of others; Creates barriers to progress without providing workable alternatives;

☐ Maintains strong positive relationships with other team members; Accepts direction

☐ Maintains strong positive relationships with other team members; Accepts direction very positively; Cooperative, demonstrating tact, courtesy and effectiveness in dealing with others; Willing to negotiate and/or compromise where appropriate;

☐ Maintains strong positive relationships with other team members. Accepts direction very positively; Cooperative, demonstrating tact, courtesy and effectiveness in dealing with others; Willing to negotiate and/or compromise where appropriate; Recognises the contributions of others; Actively contributes in assessing progress and providing workable solutions and removing barriers;

11 Leadership

☐ Avoids taking responsibility; Tends to treat some colleagues more favorably than others; Unable to effectively manage staff and/ or resources; Fails to motivate others; Fails to consider the views of others; Can react inappropriately when opinions are challenged;

☐ Generally takes responsibility appropriate to stage / experience; Treats all colleagues fairly;

☐ Willingly takes responsibility; Treats all colleagues fairly; Effectively manages staff and/ or resources; Motivates others;

☐ Willingly takes responsibility; Treats all colleagues fairly; Effectively manages staff and/ or resources; Motivates others; Welcomes and appreciates the contributions of others; Works in a harmonious manner and has highly effective working relationships;

SCHOLAR AND TEACHER

12 Learning

☐ Demonstrates little evidence of learning from experience; Does not respond positively to feedback; Struggles to recognise or acknowledge own strength and weaknesses; Does not always prepare effectively for new challenges; Reluctant to critically evaluate own performance; Fails to seek out and/ or take advantage of learning opportunities;

☐ Learns from experience; Generally accepts and acts on feedback;

☐ Learns from experience; Accepts and acts on feedback; Shows insight into own strengths and weaknesses; Conscientiously and effectively prepares for every new challenge;

☐ Recognises and uses activities as learning experiences; Accepts and acts on feedback; Shows insight into own strengths and weaknesses; Conscientiously and effectively prepares for every new challenge; Accurately critically evaluates own performance; Seeks out and takes advantage of extra learning opportunities;

13 Teaching

☐ Avoids sharing knowledge / skills with others; Teaching is poorly prepared and/ or poorly delivered; Fails to recognise/ take opportunities to provide encouragement and support; Rarely gives feedback to juniors; Feedback given in a manner that is not timely and/ or constructive; Unable to adapt teaching to the needs / questions of the individual or group;

☐ Shares knowledge / skills with others; Delivers effective teaching to relevant staff

☐ Willingly shares knowledge / skills with others; Delivers effective teaching to relevant staff; Offers encouragement and support; Gives feedback to juniors as appropriate;

☐ Willingly shares knowledge / skills with others; Delivers effective teaching to relevant staff; Offers encouragement and support; Gives feedback to juniors as appropriate; Feedback to juniors is constructive and timely; Adapts teaching to needs/ questions of individual or group;

PROFESSIONALISM

14 Professional Integrity

☐ Fails to appropriately prioritise patient(s) needs before their own; Not always open and honest; Does not always recognise professional boundaries; Not always respectful to colleagues and patients; Sometimes late and/ or unreliable – poor time management skills; Tries to justify or find excuses for errors - blames others;

☐ Appropriately prioritises patient(s) needs before their own; Open and honest;

☐ Appropriately prioritises patient(s) needs before their own; Open and honest; Recognises professional boundaries and acts appropriately; Generally respectful to colleagues and patients;

☐ Appropriately prioritises patient(s) needs before their own; Open and honest; Recognises professional boundaries and acts appropriately; Always respectful to colleagues and patients; Reliable and punctual; Acknowledges errors or misunderstandings and takes responsibility for own actions;

15 Legal, ethical & political awareness

☐ Does not always behave in an ethical and responsible manner; Sometimes does not conform to the legal aspects of informed consent and confidentiality; Does not always adhere to national / state / hospital regulations; Poor knowledge of ethical / legal implications of actions; Inadequate recognition of the medico-legal aspects of everyday practice; May overlook ethical / legal requirements under stress;

☐ Consistently behaves ethically and responsibly; Acts in accordance with the ethical / legal aspects of informed consent and confidentiality;

☐ Consistently behaves ethically and responsibly; Acts in accordance with the ethical / legal aspects of informed consent and confidentiality; Always adheres to national / state / hospital regulations; Aware of ethical / legal implications of actions;

☐ Consistently behaves ethically and responsibly; Acts in accordance with the ethical / legal aspects of informed consent and confidentiality; Always adheres to national / state / hospital regulations; Aware of ethical / legal implications on actions; Recognises the medico-legal aspects of everyday practice; Able to identify and comply with ethical expectations under stress;

16 Personal attributes

☐ Lacks enthusiasm/ commitment for the job; Lacks respect / empathy for others; Lacks stamina to cope with the physical demands of the job; Shows a negative attitude to some aspects of their work; Waits to be directed - resulting in the need for frequent supervision; 'Disappears' when problems arise;

☐ Enthusiastic / committed to the job; Treats others with some sensitivity and understanding

☐ Enthusiastic / committed to the job; Treats others with sensitivity and understanding; Stamina to cope with the physical demands of the job; Maintains a positive attitude to all aspects of their work;

☐ Enthusiastic / committed to the job; Treats others with sensitivity and understanding; Stamina to cope with the physical demands of the job; Maintains a positive attitude to all aspects of their work; Takes personal responsibility for completing tasks without prompting; Shows initiative;

APPENDIX E – General Surgery work based assessment forms

Example DOPS, MiniCEX and End of Term assessment forms are presented.

1. DOPS form
2. MiniCEX form
3. End of Term Assessment form

First Name:________________**Surname**: ____________________**Id Number**:________

Date: ______________________**Assessor Name** (completing form) ______________

Setting: ☐ Theatre ☐ ICU ☐ ED ☐ Other____________

Procedure____

Type: ☐ Major ☐ Minor Number of times performed____

Difficulty: ☐ Easier than usual ☐ Average ☐ More difficult than usual

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

Overall Score	Significant Improvement Required	Some improvement Required	Competent
Overall performance during encounter			

Suggestions for development:
Other comments:
Agreed action:

Trainee Signature: ______________________________

Assessor Signature ______________________________

First Name:_______________Surname: _________________Id Number:_________

Date: _____________________Assessor Name (completing form) ______________ __________________________

Setting: ☐ Ward/ICU ☐ OPD ☐ ED ☐ Other

Type: ☐ New case ☐ Follow-up

Focus: ☐ History ☐ Phys Ex ☐ Diagnosis ☐ Management ☐ Explanation

Complexity: ☐ Low ☐ Average ☐ High

Please assess and mark the following areas:	**Below expectations for level of training**	**Borderline**	**Meets expectations**	**Above expectations for level of training**	**Not observed Or not applicable**
1. History taking					
2. Physical Examination					
3. Communicates to patient (and family) about diagnosis, management, and potentialities to encourage their participation in informed decision making					
4. Adjusts the way they communicate with patients for cultural and linguistic differences and emotional status					
5. Recognises what constitutes 'bad news' for patients (and their family) and communicates accordingly					
6. Recognises the symptoms of,					

	accurately diagnoses, and manages common problems					
7.	Professionalism					
8.	Organisation / Efficiency					

Overall Score	**Significant Improvement Required**	**Some improvement Required**	**Competent**
Overall performance during encounter			

Suggestions for development:

Other comments:

Agreed action:

Trainee Signature: ____________________

Assessor Signature ____________________

ROYAL AUSTRALASIAN COLLEGE OF SURGEONS
BOARD IN GENERAL SURGERY

END OF TERM IN TRAINING ASSESSMENT FORM

TRAINING PERIOD	FROM: / /	TO: / /
NAME OF TRAINEE	PROBATIONARY TERM YES / NO	
No. DAYS ABSENT		REASON (eg. holiday/exam/study/illness):
HOSPITAL	(CODE: __ __ / __ __ / __ __)	
NAME OF UNIT		No. SURGEONS ON UNIT:
SURGICAL SUPERVISOR		
NAME & POSITION OF MEMBERS OF UNIT CONSULTED FOR THIS ASSESSMENT (eg. consultants, nurses, allied health) **NB: Input is required from ALL surgeons on the Unit to reach consensus in the assessment.**		

The Regional RACS Office must receive completed assessment forms no later than two weeks after the end of the training rotation. **Failure to sign and submit these forms within two weeks will result in the term not being accredited and the immediate commencement of Probationary Training.**

Notes to Surgical Supervisor on completing this form

- The Surgical Supervisor on the Unit, or a delegate, must seek the input of ALL consultant members of the Unit to reach **consensus in the assessment** of each of the competencies listed on the form. This might best be achieved at a face-to-face meeting of all consultants. Other persons who have had contact with the Trainee may also be approached to contribute to the assessment.
- The competencies listed in the 'Competent' column are those which have been identified as being required of all Trainees prior to graduation. Assessors are to categorise each Trainee's performance against each specified competence and against one of the four descriptors taking into account the Trainee's level of training.

  N - Not Competent– is lacking in competence in the designated area or is unsafe

B - Borderline – not yet competent, requires additional time, experience and/or additional training to improve;

C - Competent correctly demonstrates required competence - meets expected standard;

E - Excellent – consistently demonstrates an unusually high level of performance

- It is expected that the majority of Trainees will fall in the 'competent' category for most competencies. Assessors are asked to write in the right hand column the letter **N, B, C, E** that best reflects the Trainee's performance during the training period for each specified competency.
  - Although the assessment form may be filled out in the absence of the Trainee, the Supervisor must subsequently meet with the Trainee to discuss the assessment and to review the logbook data. Following this, the Trainee is required to sign the form and forward it together with the logbook summary to the Regional RACS Office. Both forms must be returned within 2 weeks of the end of term date. The Supervisor is advised to retain a copy of the assessment for future reference.

Notes on the responsibilities of Surgical Supervisors in managing Trainees

- Surgical Supervisors play a crucial role in the continuing formative assessment of trainees. It is important that care and attention be given to Trainee's performance of the identified competencies throughout their training.
- If a Supervisor is concerned about a trainee they are advised to record these concerns at an early stage and to ensure that both major and minor incidents are contemporaneously recorded so that any emerging pattern may be identified.
- Surgical Supervisors are obliged to inform a Trainee at an early stage of any concerns they might have. Supervisors should discuss their concerns with the Trainee in a matter-of-fact and confidential manner, and record the outcome of any discussions or interviews they might conduct.
- The outcome of such discussions or interviews should be a written plan of action to remedy the identified area(s) of concern, signed by both the Supervisor and Trainee.
- If the Trainee does not participate in any discussion/interview/plan of action in a timely fashion the Supervisor must convey their concerns in writing to the Trainee and to the Chairman of the Regional Board in their State/Country.

Notes on the responsibilities of Trainees in participating in end of term assessment and logbook review

a) It is the Trainee's responsibility to participate in the assessment process and to have the assessment form completed on time.

b) The Trainee must arrange to meet with the Surgical Supervisor to discuss the assessment and to have the logbook data reviewed. Sufficient notice must be given to allow all consultants on the Unit to meet and discuss the assessment prior to the Trainee and Supervisor meeting. If the Surgical Supervisor is to be on leave during this time, arrangements should be made to complete the form at an earlier stage

c) The Trainee must sign and return the form and logbook summary to the **Regional RACS Office** no later than **two weeks** after the end of term date. Please see the last page for contact details of the Regional RACS Offices.

d) Non-submission of a signed form within the two-week time frame will result in automatic **PROBATION** for a minimum period of 6 months and possible non-accreditation of the term. Trainees are required to **retain a copy** of this form in their portfolio records.

Notes on probationary training

- If the end of term assessment is rated 'unsatisfactory', in accordance with the General Surgery Regulations, the Trainee is **immediately placed on probationary training** for a minimum of 6 months, and pending further review by the Regional Subcommittee of the Board in General Surgery.
- Should a Trainee's overall performance be rated 'unsatisfactory' at the end of term, while on probationary training, this will constitute grounds for considering dismissal, in accordance with the College's Dismissal Policy.
- Regulations and policies relating to probationary training and dismissal are available on the College website.

- **SPECIALIST TRAINING IN GENERAL SURGERY**

Not Competent (N)	Borderline (B)	Competent (C)	Excellent (E)	Rating
MEDICAL EXPERTISE – access and apply relevant knowledge to clinical practice				
Poor knowledge base Significant deficiencies or poor perspective Allows deficiencies to persist	Needs direction to study Struggles to correctly/ accurately apply scientific knowledge to patient care	Maintains currency of knowledge Applies scientific knowledge to patient care Reads appropriately, asks for information and follows-up Recognises and solves real-life problems	Outstanding knowledge Knows common areas in depth Aware of the unusual Excellent application of knowledge in clinical situation	

TECHNICAL EXPERTISE – safely and effectively perform appropriate surgical procedures				
Fails to acquire appropriate skills despite repeated instruction/ practice. Too hasty or too slow. Rough with tissue.	Is inconsistent in retaining procedural knowledge/ skills Lacks attention to detail. Hesitant.	Consistently demonstrates acquisition, practice and retention of sound procedural knowledge, surgical skills and techniques for level of training	Excellent and SPECIALIST abilities in procedures and techniques Excellent pre-operative preparation	
Poor manipulative skills Poor hand/eye coordination	Slow in learning new skills Lapses in dexterity	Demonstrates manual dexterity required to carry out procedures Good hand/eye coordination	Outstanding technician Fluent and always in control Meticulous	
Unable to adapt skills and techniques	Ongoing weaknesses Struggles to adapt skills to different contexts	Adapts their skills in the context of each patient—each procedure	Extremely good at adapting skills for varying operative situations Excellent surgical judgement	
Lacks enthusiasm and/or initiative to participate and/or learn	Fails to improve skills and/or learn from experience	Maintains skills Effective in learning new skills	Seeks opportunities to learn new skills.	
Lacks care and diligence in approach 'Near enough is good enough'	Requires close supervision	Approaches and carries out procedures with due attention to safety of patient, self, and others	Outstanding clinician Constantly aware and responds to patient, self and team members	
As surgical assistant fails to follow operation	Has lapses of concentration	Follows the operation with guidance from the operator	Anticipates the needs of the operator & responds accordingly	
Ignores/fails to follow up problematic performance Little recognition of deficiencies in skills or techniques	Occasionally acknowledges/ follows up on problematic performance Ignores feedback	Consistently analyses their own clinical performance for continuous improvement Learns from feedback from others	Accurate in self-appraisal, excellent insight Seeks and accepts criticism & responds appropriately Aware of own skill limitations	

JUDGEMENT – clinical decision making/organise diagnostic testing, imaging and consultation as needed				
Incomplete or inaccurate Poor basic skills	Hesitant or inconsiderate of patient Lacks attention to detail.	Takes a history, performs an examination, and arrives at a well-reasoned diagnosis Efficiently and effectively examines the patient	Precise, thorough and perceptive	
Incomplete/inaccurate recognition of significant symptoms Significant errors/ omissions in diagnosis Frequent inaccuracies history, signs or diagnosis	Poor presentation/ discussion of clinical cases Occasional inaccuracies in diagnosis Sometimes confuses priorities	Recognises symptoms, accurately diagnose, and manages common disorders Differentiates those conditions amenable to operative and non-operative treatment Concise and correct on clinical details Arrives at appropriate conclusions in case presentations	Accurate and efficient Considers a wide range of symptoms and factors Insightful perspective in case discussions	
Inadequate or Inappropriate, poor selection and/or interpretation Disregards patient's needs or circumstances	Unable to appropriately justify use of selected investigations Occasional errors	Selects appropriate investigative tools and monitoring techniques cost-effectively Appraises and interprets results of investigations against patient's needs in the planning of treatment	Always selects optimal investigations Excellent interpretation Safe, efficient and cost effective approach to use of investigations	

	in interpretation that could lead to patient problems Disregards system needs	Critically evaluates the advantages and disadvantages of different investigative modalities		
Unable to make a decision Unable to suggest alternative interpretations	Some suggested alternatives are inappropriate Ignores data that does not fit interpretation Presentation unclear, disorganised	Formulates a differential diagnosis based on investigative findings Evaluates the significance of data Indicates appropriate alternatives in the process of interpreting investigations and in decision making Clear & concise presentation of findings	Precise, well organised, thorough, systematic, focused o Presentation of findings o Indicates relevant alternatives o Decisions based on data	

Not Competent (N)	Borderline (B)	Competent (C)	Excellent (E)	Rating
JUDGEMENT – *continued*				
Poor record keeping 3 incomplete, disorganised, irrelevant, illegible 4 not up-to date	Records difficult for others to follow	Contemporaneously maintains accurate and complete clinical records Precise and focused Complies with required organisational structure	Perceptive of relevant information / data for documentation Records very easily accessible	
Disinterested or indifferent approach to patients Fails to grasp significance or respond accordingly	Culturally incompetent Ignores/overlooks some patient's needs	Manages patients in ways that demonstrate sensitivity to their physical, social, cultural, and psychological needs Considers all issues relevant to the patient	Excellent and highly developed ability to manage & interact with patients and to anticipate and/or respond to their needs	
Copes poorly in situations of stress and/or complexity Under or over reacts	Can show signs of stress when managing trauma patients	Effectively manages the care of patients with trauma including multiple system trauma Maintains controlled approach & demonstrates sound judgement during times of stress/complexity	Anticipates possible risks and/or complications In stressful situations always maintains orderly approach and demonstrates sound judgment	
Inadequate planning Inadequate involvement in pre & post-operative care Fails to grasp significance of symptoms or respond accordingly	Slow to anticipate/ manage complications Slow to call for assistance Under estimates complexity and/or risk factors	Plans, and where necessary implements a risk management plan. Conscientious and reliable follow-up Effectively manage complications—operative procedures & underlying disease process Identifies and manages risk Manages complexity and uncertainty	Outstanding clinician who o anticipates possible risks/complications o identifies problems early o follows-up meticulously o coordinates and uses other personnel effectively	

Not Competent (N)	Borderline (B)	Competent (C)	Excellent (E)	Rating
COMMUNICATION – communicate effectively				
Disliked by patients because of poor interpersonal skills Bad listener Poor communicator Increases patient anxieties Patients remain confused or unclear and/or unable to follow instructions	Limited discussion with patients around issues of informed consent and/or treatment options	Trusted by patients. Listens well Communicates with patients (and family) about procedures, potentialities, and risks associated with surgery in ways that encourage their participation in informed decision making Communicates with patients (and family) the treatment options, potentials, complications, and risks associated with all treatment modalities Recognises 'bad news' for patients and relatives & modifies communicates	Possesses excellent interpersonal skills Develops excellent rapport with patients & team members Inspires confidence Patients delighted to be looked after by this trainee Demonstrates empathy appropriately	
Unaware of patient's needs Unable to communicate under varying conditions/situations	Limited perception of patient's perspective or communication needs	Appropriately adjusts the way they communicate with patients & relatives to accommodate cultural and linguistic differences and emotional status	Always interacts effectively with patients according to their social & health needs	

COLLABORATION - work in collaboration with members of an interdisciplinary team where appropriate				
Refuses to facilitate team function Does not acknowledge the contributions of others May undermine team members or function	Poor relationship with peers and other professionals Reluctant to offer assistance to other team members	Good rapport with nursing and other medical staff. Willing to help Employs a consultative approach with colleagues and other professionals Communicates effectively with and co-ordinate surgical teams to achieve an optimal surgical environment	Always willing to help even if personally inconvenient Excellent working relationship with other professionals Always supports colleagues and junior staff	
Causes disruption/problems Fails to recognise own disruptive behaviour	Ignores or fails to acknowledge misunderstandings	Initiates the resolution of misunderstandings or disputes with peers, colleagues, and others	Effectively diffuses any problems in the surgical team	
Reluctant/unable to work as a multi-discipline team member Self-focused Unreliable Fails to seek assistance with issues of patient care Ignores or is unaware of their own limitations	Lacks understanding of contributions of other professionals to patient care Works effectively with some team members but not others Slow in referring patients to other professionals	Respectful of & appreciates different kinds of knowledge and expertise which contribute to effective functioning of a clinical team Develops a patient care plan in collaboration with members of an interdisciplinary team Collaborates with other professionals in the selection/ use of various treatments assessing the effectiveness of options Recognises and facilitates referral of patients to other professionals	Excellent team member Extremely knowledgeable about the contribution of different fields of care Aware of and seeks the contribution of different fields and refers patients in a timely and appropriate manner	

Not Competent (N)	Borderline (B)	Competent (C)	Excellent (E)	Rating
MANAGEMENT and LEADERSHIP – effectively use resources to balance patient care and system demands				
Unaware of management constraints and/or expectations Reluctant to take on any management responsibility Wasteful of resources	Lacks insight into the impact of system demands Poor interaction with and/or supervision and management of junior medical staff	Identifies and differentiates between resources of the health care delivery system and individual patient needs. Effectively assesses and manages systemic risk factors Applies a wide range of information to prioritise needs and demands Directs and supervises junior medical staff effectively	Willing to contribute to health services management Uses resources very effectively for patient care balanced with patient need Excellent role model for junior medical staff, all ways offers support for junior medical staff	

HEALTH ADVOCACY				
Ignores/jeopardises own or colleagues health or well-being	Poor care of own health	Promotes health maintenance of colleagues Looks after own health	Maintains high level of fitness and encourages others	
Takes little interest in patient health beyond surgery	Limited knowledge of causal issues relating to patient health	Advocates patient health Discusses causal health issues with patient	Very knowledgeable and active in advocating patient health including preventative measures	

SCHOLAR and TEACHER – recognise the value of knowledge and research and its application to clinical practice				
Little evidence of reading texts or journals Needs direction to study	Reading of research /texts is undirected Has difficulty applying knowledge to practice	Assumes responsibility for own learning Draws on different kinds of knowledge in order to weigh up patient's problems- context, issues, needs & consequences Critically appraises new trends in General Surgery	Always keen to discover new knowledge Takes extra courses & learning opportunities	
Avoids teaching if possible. Poorly prepared, poorly delivered	Ineffective as a teacher	Facilitates the learning of others Competent and well prepared in teaching others	Enthusiastic/inspiring teacher Logical and clear Excellent teaching skills	

PROFESSIONALISM – appreciate the ethical issues associated with General Surgery				
Behaviour inconsistent with ethical ideals	Little knowledge / interest in ethical or medico-legal issues	Consistently applies ethical principles Identifies ethical expectations that impinge on common medico-legal issues	Highly conscientious Anticipates areas where medico-legal issues may arise	
Late, idle, unreliable, forgetful Off-loads work onto others	Occasionally difficult to contact or leaves tasks incomplete	Acts responsibly Dependable, conscientious Always completes tasks	Applies self beyond the 'call of duty'	
Copes poorly under stress 'Disappears' when problems arise	Pays little regard to clinical audit	Regularly participates in clinical audit Willing to undergo close scrutiny Responds appropriately to stress	Anticipates and remains efficient "when the going gets tough" Seems to thrive on pressure	
Has problems acknowledging/ recognising mistakes Unable to accept criticism	Only accepts criticism from some	Acknowledges & learns from mistakes Accountable for own decisions/actions Recognises & acknowledges their limits	Prompt response to criticism marked improvement and positive change	
Has inaccurate view of own performance	Over confident	Employs a critically reflective approach	Has great insight into their level of performance	

ESSENTIAL CRITERIA	UNSATISFACTORY =U	SATISFACTORY =S	Please Write a 'U' or 'S' below for each criteria.
Communication	Bad listener and communicator. Disliked by patients and/or nursing staff. Increases patient anxieties.	Listens well, explains well. Trusted by the patient and the nursing staff.	☐
Co-operation	Refuses to help out. Poor relationship with peers and nursing staff.	Good rapport with nursing and other medical staff. Willing to help. A team player.	☐
Self-motivation	Idle, lacking in any work enthusiasm. Behind with letters or summaries.	Hard-working, keen to learn, self organises waiting list.	☐
Work Ethic	Poor time management. Forgets to do things. Unreliable. Does not heed advice.	Dependable. Efficient in use of his/her time. Completes tasks and anticipates well.	☐
Ability to Manage Stress	Copes poorly. "Disappears" when problems arise. May show aggression towards junior medical or nursing staff.	Responds appropriately, seeks help when needed, Copes very well. Always relaxed in a crisis. Never angry nor aggressive.	☐
Honesty	Lies to cover defects in work. Does not report information correctly. Covers up errors or blames others for problems. Untrustworthy.	Honest. Admits mistakes. Trustworthy.	☐
Empathy	Relates poorly to patients and families. Arrogant.	Relates to patients and families in an appropriate manner.	☐
Teamwork	Fights with nursing staff or complaints frequently received from nursing staff about the applicant. Does not work well with junior staff or peers.	Works well with medical staff. Regarded as a team player by nursing staff. Well respected by peers and junior medical staff.	☐
Insight/Self Awareness	Lacks insight into own poor performance. Fails to take action or advice to improve performance. Denies there is an issue.	Demonstrates insight into own performance. Addresses issues when advised. Self critical and incisive.	☐

- **PLEASE NOTE: The Board considers satisfactory grades in the above non-technical criteria essential for a surgical career. A discussion with the Director of Medical Services may be necessary to gain knowledge of any staff or patient complaints received. The receipt of a 'U' in any of the above categories will result in immediate Probation for this trainee. If the trainee is already on Probation, their continuation in the training program will be reviewed.**

RESEARCH ACTIVITIES DURING CURRENT TERM: (please circle appropriate statement for each research area			
Continuing research	No current project	Research project in progress	Research project completed
Publications	No current project	Project being prepared for submission for publication	Article(s) accepted for publication and/or published

WAS A MID-TERM ASSESSMENT COMPLETED? (please circle) a) Was remedial training required? (If YES please attach copy of remedial plan) b) Has there been significant improvement as a result of remediation?	YES YES YES	NO NO NO
HAS THE TRAINEE BEEN RATED LESS THAN 'COMPETENT' IN ANY AREAS? **(please circle – if YES, this must correlate with ratings given on the form)** 1. Has each of the areas been discussed with the Trainee?	YES YES	NO NO
PLEASE PROVIDE FURTHER INFORMATION ON THE AREAS RATED LESS THAN 'COMPETENT' (if insufficient space please attach separate document).		

Note: Details of area(s) of less than competent performance must be fully documented and attached to this assessment form in addition to copies of counselling session minutes/notes from performance related discussions or meetings that have been held in the hospital

RATING OF LOGBOOK DATA (please check major cases, primary operator rates and endoscopy numbers)

Satisfactory ☐

Unsatisfactory ☐

OVERALL RATING OF TRAINING ROTATION at END OF TERM

Satisfactory This training period will be accredited towards SET. ☐

Unsatisfactory This training period will NOT be accredited towards SET. The Trainee will immediately commence on probationary training pending a review by the Regional Chair. ☐

<u>**SIGNATURE OF SURGICAL SUPERVISOR**</u> **I hereby verify that all consultants on the Unit have contributed to this assessment and that the assessment and logbook data has been discussed with the Trainee.**		
Name:	Signature:	Date:

SIGNATURE OF TRAINEE		
I have sighted the assessment on this form:	(please circle)	**YES / NO**
I have discussed the assessment with my Supervisor:	(please circle)	**YES / NO**
I agree with the assessment on this form:	(please circle)	**YES / NO**
Name:	Signature:	Date:

The Regional RACS Office must receive completed assessment forms no later than two weeks after the end of the training rotation. **Failure to sign and submit these forms within two weeks will result in the term not being accredited and the immediate commencement of Probationary Training.**

NSW & ACT Trainees send forms to:
NSW Regional Office
177A Albion Street
SURRY HILLS NSW 2010
Phone: + 61 2 9331 3933
Fax: + 61 2 9331 3145
Email: college.nsw@surgeons.org

QLD Trainees send forms to:
QLD Regional Office
50 Water Street
SPRING HILL QLD 4004
Phone: + 61 7 3835 8600
Fax: + 61 7 3832 5001
Email: college.qld@surgeons.org

SA Trainees send forms to:
SA Regional Office
51-54 Palmer Place
NORTH ADELAIDE SA 5006
Phone: + 61 8 8239 1000
Fax: + 61 8 8267 3069
Email: college.sa@surgeons.org

VIC & TAS Trainees send forms to:
VIC Regional Office
College of Surgeons Gardens
Spring Street
MELBOURNE VIC 3000
Phone: + 61 3 9249 1255
Fax: + 61 3 9249 1256
Email: college.vic@surgeons.org
college.nz@surgeons.org

WA Trainees send forms to:
Western Australian Regional Office
M307, University of Western Australia
35 Stirling Highway
NEDLANDS WA 6009
Phone: +61 8 6488 8699
Fax: + 61 8 6488 8698
Email: college.wa@surgeons.org

NZ Trainees send forms to:
New Zealand Regional Office
Elliot House
43 Kent Terrace
WELLINGTON SOUTH NZ
Phone: + 64 4 385 8047
Fax: + 64 4 385 8873
Email:

For NT Trainees - please send forms to the Regional Office responsible for the Trainee's hospital position

APPENDIX F – Flowcharts of RACS selection and assessment processes

Processes that applied to SET for the 2008. 2009, and 2010 selection years