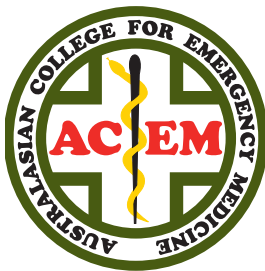




Expert Advisory Group on Discrimination Final Report to the ACEM Board

October 2017



Australasian College for Emergency Medicine

34 Jeffcott Street
West Melbourne VIC 3003

t +61 3 9320 0444
f +61 3 9320 0400
e admin@acem.org.au
w acem.org.au

Report to the Australasian College for Emergency Medicine on Discrimination in College Examinations

October 2017

Dr Helen Szoke	Chair, External Appointee
Professor Ron Paterson	Deputy Chair, External Appointee
Professor Kichu Nair AM	External Appointee
Dr Simon Judkins	FACEM; President-Elect
Dr Yusuf Nagree	FACEM; Chair, Council of Advocacy, Practice and Partnerships
Dr Mahesh Gangadharaiah	Recently qualified FACEM
Dr Hussain Kadim	FACEM Training Program trainee
Dr Danika Thiemt	FACEM Training Program trainee

Table of Contents

Table of Contents	iv
Index of Tables	vi
Index of Figures	vi
Chair Foreword	1
Executive Summary	2
1 Background	5
Fellowship Examination (pre-2015)	7
2 The College's Examination and Assessment Processes	7
Fellowship Examination (2015-)	9
Standard Setting and the Borderline Regression Method	10
Assessment of Curriculum Domains and Global Rating	11
Examination Governance	11
Court of Examiners	12
Examination Candidate Data	12
Examination Improvement Measures	13
Special Consideration, Reconsideration, Review and Appeal Processes and Examinations	14
Examination Attempts	14
Workplace-Based Assessments	14
3 The Complaint	16
The Affected Cohort	16
Particulars of the Complaint	17
Principal Remedy Sought by Complainants	17
Psychometric Analysis of the 2016.2 OSCE Data	17
Whether Unlawful Discrimination has Occurred	18
The EAG's Findings Relevant to Complaint	19
<i>The domain complaint and the examiner complaint</i>	19
<i>The results complaint</i>	20
<i>Denial of procedural fairness</i>	20
Recommendations on Appropriate Remedies for the Complainants	21
<i>Remedies requested by complainants</i>	21
<i>EAG's response to remedies requested by complainants</i>	21
<i>EAG's recommended remedies for the complainants</i>	22

Submissions from Trainees and Examination Candidates	23
<i>Subjectivity and scoring</i>	23
<i>Adequacy of preparation and training</i>	23
4 Submissions	23
<i>Inadequacy of feedback</i>	24
<i>Conduct of examiners, confederates and actors</i>	24
<i>Prevalence of stereotypes</i>	24
<i>Personal toll</i>	25
Submissions from FACEMs and DENTs	25
<i>Adequacy of preparation and training</i>	25
<i>Inadequacy of feedback in training</i>	25
<i>Inadequacy of feedback in Fellowship Examinations</i>	25
<i>Reasons for differential results</i>	26
Submissions from Examiners	26
<i>Concerns regarding the interim report/interim findings</i>	26
<i>Examination process</i>	27
<i>Adequacy of preparation and training</i>	27
<i>Reasons for differential results</i>	27
A Review of Literature on Success Rates and Examiner Bias in High-Stakes Clinical Examinations	29
5 Analysis of Differential Pass Rates	29
The College's Perspective on the Variability in Pass Rates	32
Bias in Examinations	37
<i>Patent racial bias</i>	37
<i>Unconscious bias</i>	37
<i>Impact of stereotypes</i>	37
6 Analysis of Key Issues Arising in the Review	37
Bias and the Correlation with Diversity	38
Accounting for Unconscious Bias	39
Transparency and Accountability of OSCEs	41
<i>Transition to new format</i>	41
<i>Feedback provided to unsuccessful candidates</i>	41
Associated College Processes	42
<i>WBAs and ITAs</i>	42
<i>The College's complaints process</i>	42
Culture of the College	42
7 Findings	44
8 Recommendations	46
Recommendations Dealing with Legacy Issues	46

Recommendations Aimed at Continuous Improvement of the Examination Processes	46
<i>Conduct of the examination</i>	47
<i>Support for examiners</i>	47
<i>Trainee preparation</i>	47
<i>Examination feedback</i>	48
<i>Review of results</i>	48
<i>Associated College processes</i>	48
<i>Complaints policy</i>	48
<i>In-Training Assessments and Workplace-Based Assessments</i>	49
<i>In-training supervision</i>	49
<i>Support for trainees</i>	49
<i>Culture of the College</i>	50

Annexure 1 - Terms of Reference	51
--	-----------

Index of Tables

Table 1	Pre-2015 Fellowship Examination - SCQ section score calculation	8
Table 2	Pre-2015 Fellowship Examination - SCE section score calculation	9
Table 3	Fellowship Clinical Examination - Percentage of candidates by previous attempt	13
Table 4	Fellowship Clinical Examination - Percentage of candidates by country group	34
Table 5	Fellowship Clinical Examination - Historical measures	34
Table 6	Pass rates by country group per clinical examination	36

Index of Figures

Figure 1	Variability in OSCE pass rates between each examination conducted since 2015 for all trainees sitting and who obtained clear passes, and trainees in Group B	33
Figure 2	Pass rate trends by country group per clinical examination	35

Chair Foreword

I present the findings of the Expert Advisory Group on Discrimination to the ACEM Board for consideration. In doing so, I wish to acknowledge the complexity of the challenges posed by this process. The members of the EAG are acutely aware that the events that triggered the initial complaint and the subsequent investigation have highlighted many issues – the human cost and anguish of many trainees who are grappling with the examination process to achieve Fellowship, the commitment and dedication of examiners, who are keen to uphold clinical standards and create the best possible context for patient safety and good clinical care, the processes that have most recently been instituted by College experts who seek to build a stronger and more resilient examination process to achieve Fellowship within the College system. This complexity, and the fact that inevitably processes such as this touch on human vulnerability, mean that the recommendations contained in this report require a dedication on the part of the ACEM leadership to long term and deep reform, and a commitment to focussing not just on the architecture of the examination processes but also on the deeper issue of the culture of the College and its relationship with its members.

The recommendations in this report are outlined to respond specifically to the complainants, and then to propose changes that should be made to respective college processes to build a stronger system into the future. In doing so, the EAG has sought to recognise the improvements that are part of the quality control cycle of the College and the willingness of the College to engage with these issues in a systemic manner.

Notwithstanding this, the lessons taken from this process are clear. In the first instance, there is a need to address the legacy issues of the changes in the examination process, where many graduates – both IMGs and Australian trained – have had their lives adversely impacted as they struggle to understand the requirements to achieve Fellowship. We have made specific recommendations on how these may be addressed.

There is enormous commitment and goodwill by members of the College to contribute to the profession of Emergency Medicine. We ask that you review these findings in the spirit of greater enhancement of the College's role and contribution to Emergency Medicine and the need for all institutions to be alert to unconscious bias or systemic factors that may make it difficult for certain groups to have an equal chance to achieve their professional goals.

The challenge ahead is for the College Board to take on the leadership role to respond to these findings with a comprehensive and effective action plan, to provide redress for those who have been impacted and to implement ongoing change to build the culture and effectiveness of the College into the future. I encourage clear and ongoing communication by the College Board and transparency to your membership as you respond to these recommendations.

I take this opportunity to thank the College for the opportunity to contribute to this review. I would like to acknowledge the efforts of my fellow EAG members, which were significant, in attempting to seek the best path forward, and to the Deputy Chair, Professor Paterson, for his support. ACEM staff have demonstrated high levels of competence and professionalism which greatly enhanced the ability of the EAG to undertake its work. We have been greatly assisted by Emma Turner and her colleagues in preparing and finalising our findings.

I commend this report to the Board of the Australasian College for Emergency Medicine.



Dr Helen Szoke

Chairperson

2 October 2017

Executive Summary

The Expert Advisory Group (EAG) on Discrimination was appointed by the Australasian College for Emergency Medicine (ACEM) following a complaint by an anonymous cohort of Fellowship candidates that discrimination and racial bias had resulted in the complainants (and others) failing to pass the 2016.2 Fellowship Clinical Examination (OSCE).

At the time of its appointment, the EAG was issued with terms of reference (refer Annexure 1) to guide its role in assessing and advising on steps for the College to assess and respond to the complaint, as well as respond to the broader issues which had arisen from the complaint.

The EAG undertook to assess the complaint and associated College processes to be able to make recommendations and advise the College on its next steps. As set out in the Background to this report, the EAG sought and received submissions (both written and by oral hearing), a literature review and a statistical analysis of examination data (among a plethora of additional information) to inform it of the issues it was assessing and advising on, and to guide the EAG in making the findings and recommendations set out in this report.

The EAG notes that at the time of publishing its interim report on 9 June 2017, it held a preliminary view that the introduction of the Fellowship Objective Structured Clinical Examination (OSCE) in 2015 had the unintended effect of giving rise to racial discrimination, principally through the mechanism of unconscious bias of examiners. Since the interim report, the EAG has taken further steps through a literature review and statistical analysis of OSCE data to determine whether that preliminary view was founded.

The EAG has found that the introduction of the Fellowship OSCE in 2015 as well as a number of other factors, has given rise to the complaint and examination failure for the complainants who undertook the 2016.2 OSCE.

In this report, the EAG provides its findings, assessment and advice through recommendations to its based on its enquiries, deliberations and information elicited through submissions and hearings.

It is the EAG's opinion that the impact experienced by the complainants was due to multifactorial issues as set out in this report, including:

- The College did not appropriately manage the transition to and introduction of the new Fellowship Examinations (OSCE specifically), with concessions made by the College that the process lacked information, clarity of expectations and support for candidates preparing for and attempting the examination.
- Since the introduction of the OSCE, unsuccessful candidates have not been provided with adequate feedback to enable them to understand the basis for their marks and identify areas for future improvement and practice.
- Deficiencies in the examination process (such as a lack of 'at standard' criteria setting for the domains of communication, leadership and management, teaching, and suboptimal calibration of marking criteria) coupled with comparative low examiner diversity can give rise to the risk of subjectivity and culturally laden assessments of at standard, which may disadvantage a culturally diverse candidate group.
- Prior to 2014, the College did not have entry requirements to training thereby permitting trainees to enrol with limited registration. It has been suggested to the EAG that some unsuccessful trainees may not have been sufficiently prepared for the OSCE and Fellowship.
- The College currently permits unlimited attempts at the College's examinations. From 2018, trainees will be permitted three attempts and if they do not pass, will be considered for removal from the training program. It is possible this change may have precipitated trainees taking the examination before they are ready.
- The uncoupling of the Fellowship Examinations, and the poor reliability of the 2015.1 Written Examination, may have resulted in more trainees attempting the OSCE repeatedly in succession in the last few years.
- Demands for doctors in Emergency Departments and the competitiveness of Consultant positions may have impacted workforce supply demands and may have impacted trainee preparedness influencing premature attempts at the OSCE.

- Some submissions expressed the view that the current Workplace-Based Assessment (WBA) process may be flawed in that workforce pressures influence Directors of Emergency Medicine Training (DEMTs) and Supervisors' decisions to pass trainees during these assessments despite poor performance, resulting in these assessments being 'rubber stamped' despite trainees not being clinically competent. This reluctance to accurately mark trainees may influence trainees' perception they are ready to attempt the OSCE. The EAG however notes that the WBAs do not impose a requirement on DEMTs/Supervisors to confirm a trainee's competency to undertake Fellowship Examinations. This issue requires the College's further consideration.
- The College has conceded that inadequate supervision and training of candidates in the workplace has possibly led to candidates who were not ready and/or not competent attempting the examination or continuing in the training program.
- Training is to be completed within 12 years from the time of enrolment as a trainee. Accordingly, candidates who are not otherwise ready to undertake examinations are nevertheless taking them to attempt to finish their training within the required period or else be subject to removal from training.
- There may be a true difference in performance based upon the source of a candidate's primary medical degree; that differences in their medical training can result in some candidates not being up to the examination pass standard required due to non-comparable training and assessment methods. This issue requires the College's further consideration.

The full recommendations with respect to the complaints can be found on pages 46 - 50.

As noted in the interim report and Background to this report, the EAG was initially unable to formally engage with the complainants, through their legal representative or directly, making it difficult for the EAG to form a view on the substance of the allegations in the complaint.

Following the interim report, legal representation was changed and the EAG was successful in engaging with the complainants' legal representative as well as an individual who represented the interests of the complainants through written submissions and at a hearing before the EAG. The EAG was greatly assisted by these written and oral submissions in understanding the nature of the complaint as well as the significant personal and professional impact the 2016.2 OSCE had on the complainants. The EAG understands that the personal and professional impact experienced is not limited to the complainant group given the number and content of submissions received throughout its review.

As an integral part of the initial complaint, the complainants requested that the EAG principally recommend that the results of the 2016.2 OSCE be reviewed and statistically modified to remove any element of racial bias from the results and pass grades be awarded retrospectively as well as election to Fellowship.

It is the EAG's opinion that based upon the statistical analysis undertaken for the purpose of this review, there is no statistical evidence of bias to establish that racial bias and discrimination resulted in the significant disparity of outcomes as alleged in the complaint. Accordingly the EAG does not accept the complainant's request for remedy that a statistical modification of the 2016.2 results be undertaken.

Despite this, the EAG has made a number of recommendations in this review, which go to continuous improvement measures to be undertaken by the College. The EAG also acknowledges that since the College received the complaint, it has already undertaken continuous improvement measures that were implemented prior to the 2017.1B OSCE, which was held in July of this year.

While the EAG has not identified statistical evidence of discrimination in the 2016.2 OSCE outcome, it has found that within the College there are issues with transparency in the way the Fellowship Examinations were introduced and conducted from 2015 to 2016.2. In acknowledging the measures taken by the College to date, the EAG makes further recommendations to be considered and implemented by the College – to resolve the outstanding issues which arose due to the introduction of the OSCE as well as issues which are endemic in the College's existing training processes and broader policies and procedures applicable to all College members.

The EAG makes 12 principal recommendations for the College to consider, including:

- changes to how the Fellowship Examinations are conducted; changes to supports for examiners including examiner induction, preparation and support during the conduct of examinations;
- changes to the type of feedback candidates should receive where examiners note concerns about clinical skills and performance in examinations;

- changes to the way results for Fellowship Examinations are reviewed, so that the College introduces periodical auditing as well as flagging specific examination outcomes for automatic internal review;
- changes to current College processes including how individuals are selected for appointment to key roles to increase cultural diversity and a review of College policies and procedures;
- a review of and update to the College's complaints procedure to make it transparent and bring it up to a best-practice standard;
- change to the way In-training Assessments (ITAs) are utilised to assess trainee performance and clinical capability to continue in the training program;
- change to the way in-training supervision is used to provide trainees with proper feedback that supports trainees to understand clinical deficiencies and supports trainees to work on those deficiencies and build clinical capabilities;
- changing the way the College supports trainees who received their initial medical training in a jurisdiction other than Australia and New Zealand, particularly those trainees who received medical training in jurisdictions that do not have comparable processes to Australia and New Zealand; and
- ensuring the College takes the lead in changing processes that have resulted in the complaint. This includes taking a zero tolerance approach to discrimination, bullying and sexual harassment, and ensuring that cultural diversity is a key consideration in the appointment of new positions within the College.

These are further elaborated in Section 7 - Findings (refer pages 44 - 45).

The EAG recommends that the College devise an implementation plan for the recommendations it decides to implement so that the College's actions are auditable by its members.

As highlighted in the interim report, the EAG holds concerns for the health and wellbeing of those candidates who have been unsuccessful in the Fellowship OSCE, particularly those who have provided submissions which detail their experience and the personal toll that repeated failure has taken on them. This includes psychological, professional and financial tolls. The EAG does not wish to diminish the experiences that evident in anonymous written submissions received, which demonstrated the distress and hopelessness of some trainees, including the resignation of some trainees from the FACEM Training Program.

The EAG encourages the College to publish this report (and the associated reports on the psychometric analysis and literature review) to assist the College in rebuilding its integrity with its stakeholders.

1 Background

- 1.1 The Expert Advisory Group on Discrimination (the EAG) was established in February 2017 following an anonymous complaint made by a group of candidates, self-identified as non-Caucasian candidates (NCCs), who were unsuccessful in the 2016.2 Fellowship Clinical Examination (2016.2 OSCE). The EAG's terms of reference included that it was to assess and advise on concerns of discrimination arising from that examination. The complainants undertook their own candidate polling which indicated that the likely pass rate for NCCs was 6.8% compared to 88% for Caucasian candidates (CC), again a self-identified term. On this basis the complainants alleged the College's OSCE was racially discriminatory against NCCs and sought a reconsideration of the results. They allege that the only explanation for this differential pass rate was that the OSCE was systemically discriminatory on the basis of race. Significantly, the members of the complainant group did not want to be identified because of fear of further discrimination and victimisation.
- 1.2 The College has given the EAG a broad remit to assess discrimination in relation to the College Fellowship Examinations (written and clinical), with a view to advising the College in respect of the complaint, inquiring into discrimination in the College Fellowship Examinations and, where appropriate, recommending strategies to eliminate discrimination in the College's Fellowship Examinations.
- 1.3 The EAG was established as an independent group of individuals comprising members from within the College as well as experts from the community. It has always viewed its role as investigative and advisory, and not as a determinative body. The views and opinions expressed in this report are the collective views and opinions of this group of individuals. This report does not make legal findings or constitute legal advice. The EAG received considerable assistance from the College's Manager for Standards in its work. Legal and policy support to the process was provided by Russell Kennedy lawyers, led by Ms Emma Turner.
- 1.4 As noted above, when the complaint was received, the complainants identified the cohorts of those who had passed and failed the 2016.2 OSCE as Caucasian candidates (CCs) and non-Caucasian Candidates (NCCs) respectively. In this report, a variety of acronyms are used to identify different cohorts of candidates. Where referring to the complaint and the submissions received by the EAG, the terms CCs and NCCs are used as these were the identifiers used by the complainants themselves and the terms used predominantly throughout submissions. The EAG understands that this group are not exclusively international medical graduates, with a few Australian trained doctors identifying as southern Asian origin and in the affected cohort. As an additional part of the EAG process (identified in detail below) an analysis of the pass and fail rates of candidates was undertaken by the College. This analysis uses the term Group A to refer to Australian and New Zealand trained doctors and these whose primary medical degree was issued by a country with comparable health care services (United Kingdom, United States of America, Canada and Ireland) and Group B to refer to those whose primary medical degree was issued by any other country. The College does not collect data on whether trainees identify as Caucasian or another ethnicity, therefore as the complainant cohort is anonymous it has been challenging to define the affected cohort with precision. The term 'IMG' is also used in this report to refer to 'International Medical Graduates', a broadly comparable term to NCCs. Regardless of which characterisation is preferred, the EAG encountered difficulty accurately delineating the cohort affected by the 2016.2 OSCE.
- 1.5 The EAG was given a three-month period to inquire into the matters raised in the complaint and report to the College Board on 19 June 2017. It became clear that the initial period was insufficient for the EAG to complete the review given the complexity of the issues and the further enquiries that needed to be made. Accordingly, the EAG provided the Board with an interim report in June 2017, requesting further time to consider the issues and to conduct a literature review and further statistical analysis of the OSCE data. Further time was granted and this report will be presented to the Board on 9 October 2017.
- 1.6 The EAG received the following information to assist it:
 - 1.6.1 Written submissions from a range of stakeholders, trainees (current and former), FACEMs, Supervisors and DEMTs, and examiners.
 - 1.6.2 Further anonymous submissions provided directly to the Chair of the EAG from individuals who were too fearful to contribute directly to the EAG process for fear of retribution from, and unfair treatment by, the College and Fellows, if their opinions were voiced openly through this process.

- 1.6.3 Oral submissions from representatives of the complainants (legal and individual), staff of the College, FACEMs who are or have held supervisory or examiner roles, and representatives from the Court of Examiners.
- 1.6.4 Data from the College:
- (a) comparing the pass rates of Group A and Group B candidates' performance in OSCEs since 2015 and clinical examinations from 2010;
 - (b) comparing the Workplace Based Assessment (WBAs) performance of Group A and Group B candidates to determine whether Group B students perform more poorly in the workplace and whether the OSCE process reflects this; and
 - (c) demographics of trainees and examiners based upon the country in which they obtained their principal medical degree.
- 1.6.5 Commissioned Professor Lambert Schuwirth from Flinders University, to review the 2016.2 OSCE results for any statistical evidence of bias against IMGs and/or Group B candidates.
- 1.6.6 Commissioned a literature review by Professor Elizabeth Farmer, to conduct a simplified thematic analysis of the literature relating to:
- (a) Studies of examiner bias based upon country of original medical training;
 - (b) Performance of IMGs and trainees from Non-English Speaking Backgrounds in OSCEs; and
 - (c) Strategies suggested in the literature to assist examination bodies to address examiner bias.
- 1.6.7 The results from the College's survey on Discrimination Bullying and Sexual Harassment (DBSH) conducted as part of the College's activities to identify and inform activities to reduce discrimination. Specifically, two questions related to discrimination in assessment and the Fellowship examination process.
- 1.7 For the purposes of this report the EAG has adopted the following definitions of racism, systemic discrimination and unconscious bias:
- 1.7.1 "racism is a set of beliefs that asserts the natural superiority of one group over another on the basis of race, religion, colour, descent, national or ethnic origin, and which is used to justify unfair or differential treatment and/or social positions"¹
- 1.7.2 "systemic discrimination involves the application of beliefs, values, presumptions, structures and processes by the economic, political, social or cultural institutions of society in ways that result in differential and unfair outcomes for one or more social groups".²
- 1.7.3 "unconscious bias occurs when an individual's subconscious prejudicial beliefs or unrecognised stereotypes about individual attributes, such as ethnicity, gender, socioeconomic status, age and sexual orientation, result in an automatic and unconscious reaction and/or behaviour".³
- 1.8 The College collects data on country of issue of primary medical qualification. The College has grouped and analysed this data according to country of primary medical training from health care systems comparable with Australia – the comparison groups being Australia, New Zealand, United Kingdom, United States of America, Canada and Ireland (Group A) and all other countries (Group B).

1 Extrapolated from HREOC, *"Combating racism in Australia: a discussion paper by HREOC for the World Conference Against Racism"* (believed to be 2000); VEOHRC, *"Harnessing Diversity: Addressing racial and religious discrimination in employment"* (2008); VEOHRC *"Reporting racism"* (2013).

2 Extrapolated from Victorian Health Promotion Foundation, *"More than tolerance: Addressing discrimination affecting migrant and refugee communities"* (2007); VEOHRC, *"Systemic racism as a factor in the over-representation of Aboriginal people in the Victorian criminal justice system"* (2005); VEOHRC, *"Harnessing Diversity: Addressing racial and religious discrimination in employment"* (2008); VEOHRC *"Reporting racism"* (2013).

3 Extrapolated from Santry H., MD, MSa, Wren S., MD,b,c, *Surgical Clinics of North America*. 92(1); 137 *"The Role of Unconscious Bias in Surgical Care and Outcomes"* (2012); Australian National University, *"Unconscious bias"* (2015); McCormack, H., Kenan-Flagler Business School Executive Development, University of North Carolina, Chapel Hill, NC, *"The Real Effects of Unconscious Bias in the Workplace"* (2015).

2 The College's Examination and Assessment Processes

- 2.1 Prior to 2015, the College had a different format for the Fellowship Examination whereby the examinations were 'coupled' in that candidates were required to successfully pass the written examination before being invited to take the clinical examination. If a candidate failed the clinical examination they were required to retake the entire process from the start.
- 2.2 In 2015, the College introduced the new Fellowship Examination whereby the examinations were 'uncoupled'. Despite the continued requirement to pass the written examination before undertaking the clinical examination, if a candidate fails the clinical examination they are no longer required to also retake the written.
- 2.3 The EAG received a number of submissions regarding the differences between the pre-2015 Fellowship Examination and the current Fellowship Examination introduced in 2015. Understanding the differences in the two was central to the work of the EAG and a summary is provided for the purposes of context.

Fellowship Examination (pre-2015)

- 2.4 *Written Component of the Fellowship Examination (pre-2015)*
 - 2.4.1 The written component of the Fellowship Examination that operated until the end of 2014 consisted of three separate papers: Multiple Choice Questions (MCQs), Short Answer Questions (SAQs), and Visual Aid Questions (VAQs).
 - 2.4.2 The MCQ paper comprised 60 questions, with the pass mark set at 33 (i.e. 55%). Marking was conducted by the Education Testing Centre at the University of New South Wales.
 - 2.4.3 The SAQs were equally weighted out of ten, with candidates required to pass at least five questions and attain a score of not less than 40. Candidates were marked by two examiners, with a consensus mark recorded.
 - 2.4.4 The VAQs were equally weighted out of ten, with candidates required to pass at least five questions and attain a score of not less than 40. Candidates were marked by two examiners, with a consensus mark recorded.
- 2.5 *Clinical Component of the Fellowship Examination (pre-2015)*
 - 2.5.1 The clinical component consisted of a long case, four short cases and a set of six Structured Clinical Examinations (SCEs); it was typically held approximately 10 - 12 weeks after the written component.
 - 2.5.2 The Long Case comprised candidates having 35 minutes with one patient, a five-minute 'break', followed by 20 minutes with the examiner(s). The examiner(s) allocated a mark out of ten; this being the final mark for the long case. The focus of the time with the examiners was on meeting the defined core competencies of the clinical long case through an unstructured discussion facilitated by the examiner(s).
 - 2.5.3 The Short Cases comprised candidates having four ten minute cases, each with two patients. For each of the four short cases, examiners allocated a raw mark out of ten. The raw marks were totalled and converted to a consolidated score out of ten as set out in Table 1 (overleaf).

Table 1 Pre-2015 Fellowship Examination - SCQ section score calculation

Raw Score (/40)	Final Section Score
0	0
<11	1
11 - 12	2
13 - 14	3
15 - 18	4
19 - 22	5
23 - 26	6
27 - 29	7
30 - 31	8
32 - 33	9
>33	10

To pass the short case component, a candidate needed to pass at least two of the cases (i.e. gain a score of ≥ 5). A candidate who passed three cases and achieved a raw score of 15 - 18 was awarded a pass (five marks).

- 2.5.4 The SCE comprised six stations over one hour, with two examiners per station: one examiner assumed the role of lead examiner, while the second acted as a scribe. Candidates were marked by both examiners, with a consensus mark, out of ten, recorded. Candidates were required to pass four stations and have a total score of not less than 30 to pass the SCE component. The SCE sought to test a candidate's
- (a) ability to assess a clinical scenario and prioritise diagnostic and management endeavours;
 - (b) ability to manage complex, acute situations including resuscitation of the critically unwell patient;
 - (c) verbal and non-verbal communication skills;
 - (d) description of visual data;
 - (e) team function and leadership;
 - (f) clinical administration; and
 - (g) manual dexterity.
- 2.5.5 The SCE might involve an interaction with a trained actor to facilitate the assessment of the patient encounter and crisis management, and the examination allowed examiners to assess candidates across a range of subjects addressed in the curriculum and at consultant level. It assessed a range of skills, including problem solving, clinical judgement, communication skills, interpretation of investigations, assessment and management skills.
- 2.5.6 The assessment was undertaken by the examiners, with a single consensus global mark out of ten given at the end of each station. The marks were then converted to a consolidated score as set out in Table 2.

Table 2 Pre-2015 Fellowship Examination - SCE section score calculation

Raw Score	Number of Stations Passed	Final Section Score
0	0	0
21	1	1
23	1	2
25	2	3
27	3	4
30	4	5
33	5	6
36	5	7
39	6	8
42	6	9
45	6	10

2.5.7 The Long and Short Cases were examined on a single day; the SCE was examined on a separate day.

Fellowship Examination (2015-)

- 2.6 As noted above, the current Fellowship Examination, consists of *uncoupled* written and clinical components.
- 2.7 To be eligible to sit the Fellowship Written Examination, trainees must have satisfactorily completed and been credited with at least 12 FTE months of 'core' emergency medicine Advanced Training (net of any remediation time completed or in progress).
- 2.8 To be eligible to sit the Fellowship Clinical Examination, trainees must have: passed the Fellowship Written Examination; satisfactorily completed and been credited with at least 36 FTE months of Advanced Training time (net of any remediation time completed or in progress); and completed the research requirement.
- 2.9 Criterion-referenced standard setting is required by the Australian Medical Council (AMC) for Specialist College accreditation and all components of the College Fellowship Examinations are set to an absolute standard that denotes the required/acceptable level of competence or performance that needs to be demonstrated in order to pass a particular question or station. The standard set by ACEM is the level of a graduate Fellow/'new' FACEM (at the AT3 level outcomes of the *ACEM Curriculum Framework*).
- 2.10 *Fellowship Written Examination (2015-)*
- 2.10.1 The Fellowship Written Examination is held on a single day and consists of two papers: Select Choice Questions (SCQs); and SAQs administered separately, with a break in between.
- (a) The SCQ paper consists of MCQs and Extended Matching Questions (EMQs), with the proportion of MCQs to EMQs in each paper not fixed. The paper is delivered online via the College's e-Learning platform.
 - (b) In the SAQ paper, candidates are given clinical scenarios, followed by related questions. The paper is intended to assess the ability of the trainee to both recall knowledge, as well as the prioritisation and application of knowledge.

- 2.10.2 The College employs a modified Angoff standard setting method⁴, for both the SCQ and SAQ papers, through which a 'cut score' for each paper is achieved. One Standard Error of Measurement (SEM) is added to the estimated 'cut score' to arrive at the examination pass mark.
- 2.10.3 There is no pre-determined pass mark for the examination and there is also no specified number of questions that a candidate is required to 'pass'. Candidates are required to 'pass' each paper as determined by the standard setting process.
- 2.11 *Fellowship Clinical Examination (2015-)*
- 2.11.1 The Fellowship Clinical Examination introduced from 2015 is an Objective Structured Clinical Examination (OSCE). The OSCE comprises up to 16 clinical stations, which may include the use of standardised patients, observation stations, clinical scenarios and simulations of management of critically ill patients. Rest stations are used as needed having regard to candidate numbers.
- 2.11.2 OSCE stations are ten minutes' duration (seven minutes plus three minutes' reading time); however, 'double length' stations are also used in order to assess more complex competencies and a wider range of domains than the standard ten-minute stations.
- 2.11.3 There are either one or two examiners present in each station. Prior to the 2017:1 OSCE, where there were two examiners present, a 'consensus' mark was used to arrive at a score for the performance of each candidate in the station. The EAG was advised that this practice was discontinued from the 2017:1 OSCE and separate marks from each examiner are now used.
- 2.11.4 The OSCE is conducted across multiple days, with different candidate cohorts on, for example, days 1 to 3, compared to days 4 to 6 of an examination. Currently, different stations are used for each cohort, although earlier sittings of the OSCE involved quarantining candidates and re-using the questions. The examinations are standard set separately and treated as different examinations in terms of pass mark determination.
- 2.11.5 There is no pre-determined passing score for the examination and there is also no specified number of stations that candidates need to 'pass'. Candidates are required to reach the passing score as determined by the borderline regression standard setting method.

Standard Setting and the Borderline Regression Method

- 2.12 Standard setting is the process by which a 'pass mark' or 'cut score' in an examination is determined, with the cut score being selected as the measure reflecting a satisfactory level of performance in the examination.
- 2.13 In high-stakes examinations, a SEM is frequently applied to the cut score, increasing the probability and confidence that a candidate's score truly reflects the true score of a candidate. In high stakes testing involving consideration of public safety, the EAG was advised that pass marks are often determined through the addition of a SEM to a 'cut score' in order to increase confidence in the candidates deemed to have met the cut score.
- 2.14 The borderline regression standard setting method is undertaken after the OSCE is conducted. After adjustment for domain weightings, all trainees' raw scores for each station, and their corresponding global ratings, are regressed to a line of best fit. Individual 'cut' scores are calculated for each station, for each cohort. The 'just at standard' global rating is then used to identify the corresponding station cut score from the line of best fit. Station cut scores are then combined to arrive at the 'raw' cut score for the examination cohort. One SEM is added to arrive at the 'passing' score for the examination.
- 2.15 The EAG noted advice to the College that the OSCE "format and the standard setting procedure are well known and studied, with the Borderline Regression method referred to frequently as a 'gold standard' for setting the passing standard for such assessments"⁵

4 Through a standard setting workshop, a panel of trained standard setters collaborate to make judgements of the expected performance of a 'just at standard' candidate for each examination question. For this examination, the standard setting panel includes trainees in Advanced Training. These judgements are then combined to calculate a cut (passing) score for each paper.

5 Professor E Farmer, *Standard Setting for the Fellowship Examination of the Australasian College for Emergency Medicine – Preliminary Review*, 24 November 2016, p.12.

Assessment of Curriculum Domains and Global Rating

- 2.16 The *ACEM Curriculum Framework* comprises eight domains: medical expertise; prioritisation and decision making; communication; teamwork and collaboration; leadership and management; health advocacy; scholarship and teaching; and professionalism. All domains are examinable in the OSCE.
- 2.17 Each station typically assesses two to three domains. With effect from the 2016.2 OSCE the domains are 'weighted', with the weighting of the domains assessed in a single station established in the writing and development of that station. The EAG received information on the percentage of each domain in the total marks for the 2015.1 – 2016.2 OSCEs. Noting some of the submissions received, *communication* comprised 14% and 11% of the total marks for the two 2016.2 OSCE cohorts (i.e. days 1 - 3 and 4 - 6 respectively), whereas it constituted 32% of the total marks in the 2015.1 OSCE; 22% of the 2015.2 OSCE; and 31% and 24% of the two 2016.1 OSCE cohorts.
- 2.18 For each domain, examiners are required to indicate the single best option that represents the candidate's level of performance, with each of the options corresponding to a numerical mark of 1 - 7, respectively:
- 2.18.1 Very poor level of competence displayed;
 - 2.18.2 Well below minimum level of competence displayed;
 - 2.18.3 Below minimum level of competence displayed;
 - 2.18.4 Minimum level of competence displayed;
 - 2.18.5 Above minimum level of competence displayed;
 - 2.18.6 High level of competence displayed; and
 - 2.18.7 Very high level of competence displayed.
- 2.19 In calculating a candidate's score, the assessments given are then weighted according to the applicable domain weighting. For each domain, more detailed assessment criteria are provided on the mark sheet to inform the examiner.
- 2.20 Examiners are also required to provide an overall global rating of the candidate's performance in the entire station according to the following:
- 2.20.1 Well Below Standard;
 - 2.20.2 Below Standard;
 - 2.20.3 Just at Standard;
 - 2.20.4 Above Standard; and
 - 2.20.5 Well Above Standard.

The 'standard' referenced is that of a safely practising newly-qualified FACEM.

Examination Governance

- 2.21 ACEM has a range of entities that are involved in different aspects of the examinations associated with the FACEM Training Program:
- 2.21.1 Examinations Subcommittee;
 - 2.21.2 Court of Examiners;
 - 2.21.3 Primary Examination Working Groups;
 - 2.21.4 Fellowship Examination Working Groups; and
 - 2.21.5 Panel of Standard Setters.

- 2.22 The Court of Examiners, Examination Working Groups and the Panel of Standard Setters report to the Examinations Subcommittee, which reports to the Specialist Training and Assessment Committee (STAC) although some motions and recommendations proceed directly from the Subcommittee to the Council of Education.

Court of Examiners

- 2.23 Fellows must have three or more years' post-Fellowship experience before they can apply for appointment to the Court of Examiners, with examiners expected to participate in at least ten days of examination attendance in each calendar year across the Primary Examination Viva, the Fellowship SAQ examination and the Fellowship OSCE. The College typically calls for expressions of interest for appointment to the Court of Examiners twice per year, with ten places available per examination. The EAG was advised that this is the maximum number of individuals that can be accommodated within existing arrangements to ensure all receive the necessary training; however, an examination with two candidate cohorts (days 1 - 3 and 4 - 6), does offer the opportunity for the training of up to ten new examiners per examination cohort.
- 2.24 The EAG was further advised that when applications are reviewed, consideration is given to matters such as whether there are other examiners at the applicant's employing hospital, ensuring examiners are drawn from sites with varying levels of ACEM accreditation (for the purposes of maximum Advanced Training time) and geographical distribution.
- 2.25 *Senior Examiners* are experienced examiners who undertake the usual tasks of examiners but may take on additional responsibilities such as in-examination coordination and question writing and review. Examiners who have satisfactorily completed at least five years on the Court of Examiners are eligible to become a Senior Examiner and may apply to the Examinations Subcommittee for appointment.
- 2.26 *Peer Support Examiners* are Senior Examiners who play a role in supporting individual members of the Court of Examiners, including through the training of new examiners and provision of support to examiner peers. Senior Examiners are eligible for appointment as a Peer Support Examiner.
- 2.27 The EAG was provided with data from the College on the demographics of the Court of Examiners, with the following⁶ of particular note:
- 2.27.1 Ordinary members: 34% female and 66% male; 93% Group A and 7% Group B; duration of appointment on the Court of Examiners 0 - 9 years.
- 2.27.2 Senior Examiners: 24% female and 74% male; 98% Group A and 2% Group B; duration of appointment on the Court of Examiners 13 - 29 years.
- 2.27.3 Peer Support Examiners (a discrete subset of the Senior Examiners): 20% female and 80% male; 100% Group A; duration of appointment on the Court of Examiners 21 - 29 years.

Examination Candidate Data

- 2.28 The 'Group A', 'Group B', 'Caucasian Candidate' and 'non-Caucasian Candidate' data cohorts have been previously explained; however, prior to receipt of the originating complaint, this was not data that the College collected for analysis. The College has, since the 2016.2 OSCE, reviewed and analysed data on passing candidates by previous attempts at an examination.
- 2.29 That analysis has confirmed that 'the vast majority of candidates who pass the examinations do so within three or four attempts at the examination in question'.⁷ The EAG acknowledges that this is not unique to ACEM examinations and is supported by a significant body of literature.
- 2.30 The data received by the EAG (refer Table 3, overleaf) revealed a significant increase in the number of candidates attempting the examination for the fourth or subsequent time in the 2016.2 examination.

⁶ Data as at 14 July 2017.

⁷ ACEM Reaccreditation Submission to the Australian Medical Council and the Medical Council of New Zealand, June 2017, p.133.

Table 3 Fellowship Clinical Examination - Percentage of candidates by previous attempt

Examination	Overall examination pass rate (%)	Candidates in each category (%)			
		Number of previous attempts at clinical			
		0	1	2	≥3
2015.1-C1	76.9	90	8	2	0
2015.1-C2	70.7	76	20	4	0
2015.2	55.6	68	27	5	0
2016.1A	42.4	58	19	20	3
2016.1B	38.0	55	20	24	1
2016.2-C1	46.2	46	24	11	19
2016.2-C2	53.1	66	17	5	12
2017.1A-C1	77.6	52	20	13	15
2017.1A-C2	79.1	55	9	20	16
2017.1B	50.7	41	13	14	32

- 2.31 In considering the College's reports on the 2017.1A and 2017.1B OSCE, the College's analysis of the difference of the two cohorts of candidates noted that 15% and 16% of the two 2017.1A OSCE two cohorts (i.e. days 1 - 3 and days 4 - 6) attempted the examination for the fourth or subsequent time, compared with 32% of 2017.1B candidates.

Examination Improvement Measures

- 2.32 The EAG received information from the College in relation to a range of information on improvement measures already implemented or currently being progressed in relation to the College's examinations. These include question writing workshops and review of questions in the *item bank*; the introduction of SAQ marking centres, with examiner calibration sessions; SAQ and SCQ standard setting workshops; clarification of the definition of a 'Just at Standard' candidate and consistent use of terminology; and the use of video recording of OSCE stations for the purpose of examiner training.
- 2.33 Improvement measures that directly relate to the OSCE and which have already been implemented by ACEM include:
- a written briefing on unconscious bias having been sent to all examiners prior to the May (2017.1A) and July (2017.1B) OSCEs and incorporated into the examiner briefing given at the examination;
 - two new documents, *Information about the Fellowship Examination - Clinical (OSCE)* and *Preparing for the Fellowship Examination - Clinical (OSCE)*, designed to assist trainees and their supervisors in preparing for the OSCE and published on the College website;
 - allocation to each examination candidate for each examination of a unique identifier, rather than ACEM membership number;
 - the introduction of independent rather than consensus marking for OSCE stations with two examiners;
 - enhancements to the pre-examination workshops;
 - the introduction of a calibration session after one round of candidates have been examined;
 - modifications to the mark sheet to include a separate section for examiners to provide comments in relation to areas of concern;
 - significant changes to the written feedback provided to unsuccessful candidates; and
 - significant changes to the feedback provided to examiners.

Special Consideration, Reconsideration, Review and Appeal Processes and Examinations

- 2.34 As is required of all accredited providers of specialist training and CPD programs, the College's *Reconsideration, Review and Appeals Policy* offers three layers of redress to individuals who are dissatisfied with a College decision. Each layer of redress is based on eight grounds, set out in the policy. Section 3.2(2) of that Policy states:

"This policy is not intended to provide an avenue for any person to contest results awarded, decisions of assessors taken in or about any assessment, allegations of poor training or supervision or general grievances."

- 2.35 Notwithstanding that limitation, in the period April 2015 – May 2017, seven applications for reconsideration of an examination result were received and considered, and likewise one application for review.
- 2.36 The College's *Exceptional Circumstances and Special Consideration Policy*, is also relevant to examinations and allows for consideration of individual circumstances *prior to or at the time* of decision making (e.g. requests for special arrangements in relation to known circumstances made in advance of the date of the examination and circumstances, such as acute ill health, that arise on the day/at the examination).

Examination Attempts

- 2.37 At present and historically, there is no limit on the number of occasions on which a trainee can attempt the Primary or Fellowship examinations.
- 2.38 In December 2015, the Council of Education approved, in principle, the introduction of a maximum number of attempts for trainees and in 2016 the decision was made to limit candidates to a maximum of three attempts at each of the Primary and Fellowship Written and Clinical Examinations.
- 2.39 To afford trainees adequate notice, the limit commences from 2018. However, previous attempts at a particular examination do not count toward the maximum number allowed. That is, a trainee who has attempted the Fellowship OSCE on five occasions, but has yet to pass will, from 1 January 2018, have three attempts in which to pass the examination.
- 2.40 Trainees who fail to pass an examination within the three attempts will be considered for possible dismissal from the FACEM Training Program.

Workplace-Based Assessments

- 2.41 The FACEM Training Program utilises a suite of assessment tools that include Workplace-Based Assessments (WBAs) and the examinations.
- 2.42 *In-Training Assessments (ITAs)*
- 2.42.1 The ITA is a WBA that involves the trainee being assessed by their DEMENT/Supervisor at a point in time during their placement. ITAs are completed by the trainee's DEMENT/Supervisor, and occur every three FTE months.
- 2.42.2 The EAG was advised that having recognised some limitations of the existing ITA, the Council of Education has convened a working group to revise the ITA form to align the rating scale with the outcomes expected for the domains of the *ACEM Curriculum Framework* at the relevant stage of training.
- 2.42.3 The assessment scale currently used on ITAs asks the DEMENT to select one of the following:
- (a) Senior clinician performs; trainee assists;
 - (b) Trainee performs; senior clinician instructs;
 - (c) Trainee performs; senior clinician directly observes;
 - (d) Trainee performs; senior clinician checks prior, during, upon completion;
 - (e) Trainee performs; senior clinician checks prior and upon completion;
 - (f) Trainee performs; senior clinician checks upon completion;

- (g) Trainee performs; senior clinician available to check in-person;
 - (h) Trainee performs; senior clinician available to advise via telephone; and
 - (i) Trainee performs; senior clinician not required.
- 2.42.4 The DEMENT is also asked to select the most appropriate level of involvement of a senior clinician in order for the trainee to perform at the level of a new FACEM. That question and assessment scale is the same, regardless of whether the trainee is undertaking Provisional or Advanced Training.
- 2.42.5 For Provisional trainees, the DEMENT is also asked, based on their knowledge of the trainee at that point in time, to rate the trainee's overall readiness for ACEM Advanced Training, and to select either:
- (a) Not yet ready to progress to ACEM Advanced Training; or
 - (b) Ready to progress to ACEM Advanced Training
- 2.43 *Emergency Medicine Workplace-Based Assessments (EM-WBAs)*
- 2.43.1 EM-WBAs are requirements of Advanced Training and involve periods of observation of performance and/or discussion with a trainee in clinical practice, followed by structured feedback to the trainee and a rating of the trainee's performance during the specified period.
- 2.43.2 The EM-WBAs associated with Advanced Training comprise:
- (a) *Case-based Discussion (CbD)*: where the assessor engages the trainee in discussion of a selected case, which the trainee managed, to assess and provide feedback on the trainee's clinical reasoning and decision making.
 - (b) *Direct Observation of Procedural Skills (DOPS)*: where the trainee is directly observed whilst performing a specific clinical procedure, to assess and provide feedback on trainee performance of the procedure.
 - (c) *Mini Clinical Evaluation Exercise (Mini-CEX)*: where the trainee is directly observed while performing a focused clinical task during a specific patient encounter, to assess and provide feedback on trainee performance in the patient encounter.
 - (d) *Shift Report*: where the trainee is observed for the duration of a clinical shift, to assess and provide feedback on trainee performance during a discrete time period of clinical work.
- 2.43.3 When rating a trainee's performance in an EM-WBA for the listed components, assessors select from the same nine ratings as used on an ITA, with an additional tenth option of:
- (a) Did not observe enough to make a judgement.
- 2.44 Trainee progress through the FACEM Training Program occurs by means of review by a Regional WBA Panel at defined 'progression points'. At each progression point, the relevant Regional WBA Panel makes one of the following progression decisions: *eligible to progress or not eligible to progress*.
- 2.45 Decisions on eligibility to progress are not directly linked to eligibility to sit the Primary or Fellowship Examinations so long as the trainee has attained the required number of months of training to be eligible to sit the examination in question. Similarly, even if a trainee is currently undertaking a required period of remediation this does not preclude them from sitting the examination as long as they have otherwise met the eligibility requirement(s). Additionally, WBA Panel review does not include a decision on whether the trainee is approved or ready to sit an examination.

3 The Complaint

- 3.1 In January 2017, the College received an application for reconsideration and a complaint of discrimination from an anonymous group of Fellowship Examination Candidates who described themselves as Non-Caucasian Candidates (NCC) in relation to the 2016.2 OSCE results. Their complaint alleged that there was a considerable difference in the pass rate of Caucasian Candidates (CC) (88%) as against the complainant cohort of NCCs (6.8%).

The Affected Cohort

- 3.2 These pass rates were arrived at by the complainants following their consultation with a large pool of candidates who undertook the examination; of whom 59 candidates identified themselves as being of non-Caucasian background and only four had passed. Of note, the complainants advised that this affected cohort was diverse and consisted of candidates who were “coloured, or not white”, including:
- 3.2.1 candidates of diverse nationalities and ethnicities including India, Pakistan, Sri Lanka, Malaysia, Singapore, the Philippines, Iraq, Iran, Nigeria, South Korea, Bangladesh;
 - 3.2.2 many candidates who do speak English as their first language including those who are second and third generation Australians of non-Caucasian background; and
 - 3.2.3 many candidates who are International Medical Graduates (IMGs) having obtained their primary medical degree overseas and some who obtained their primary medical degree in Australia or from countries with comparable health care systems such as the United Kingdom.
- 3.3 Significantly, various members of this group have said that they did not want to be identified for fear of further discrimination and retribution from the College and the FACEM community. Independent members of the EAG were also approached by some of these candidates and others wanting to share their experiences of discrimination and the College anonymously.
- 3.4 The EAG acknowledges:
- 3.4.1 the personal and professional impact this process has had on the complainant group and others affected by these concerns who have come forward and contributed to this process; and
 - 3.4.2 the vulnerable position that this group is in and recognises the courage and goodwill in engaging with the EAG in an attempt to resolve their concerns and make the most of the EAG’s advisory role in relation to the College.
- 3.5 The EAG also acknowledges the steps that the College has undertaken in response to this complaint and the establishment of the EAG process, including making available free, confidential and anonymous counselling and support services to trainees should they require it.
- 3.6 The anonymity of this group has presented a challenge for the EAG in engaging with complainants. The EAG initially experienced significant difficulties coordinating a time for the complainants’ original legal representative to meet with the EAG. In July, the EAG was advised that some of the complainants had changed lawyers and the composition of the group had reformed with this change, going from 30 to 20 complainants. The new group proceeded on the basis of the original complaint and included three individuals not party to the original complaint. This reformed group of complainants was represented by Parke Lawyers, and presented further oral and written submissions to the EAG on the personal impact of failed examinations attempts, further detailing concerns about how the OSCE may have a discriminatory impact and how the impacts of the 2016.2 examination could be remediated.
- 3.7 Although the initial complaint was based upon the results of the 2016.2 OSCE, the reformed complainant group sought to extend the complaint to all Fellowship Clinical Examinations since 2015. This extension was sought on the basis that there was a consistent differential in the pass rates between CC and NCC since 2015, and a finding of discrimination in relation to 2016.2 would imply that all previous OSCEs were also affected. The EAG considered that to extend its enquiries beyond the analysis of the 2016.2 results would be outside of its terms of reference. Thus the EAG findings on the complaint relate solely to the 2016.2 OSCE.

Particulars of the Complaint

- 3.8 The basis for the complainants' contention of discrimination rests in the differential pass rates of CCs compared to NCCs, which, on the complainants' data suggests that a CC is 13 times more likely than an NCC to have passed the 2016.2 OSCE. The complainants say this is statistically significant when at least 25% of the examination candidates were not "white". The complainants allege that the statistical likelihood of so few "coloured" candidates passing the examination compared to "white" candidates, due to factors other than race, must be so small to be statistically insignificant. Particularly, in circumstances where they have undergone advanced training, been assessed by FACEMs as performing at or near the level of a specialist emergency medicine physician, and have been working in Australian EDs over many years. The complainants allege that these assumptions support an inference of unlawful racial discrimination in breach of the Racial Discrimination Act 1975 (Cth).
- 3.9 The complainants have particularised their discrimination complaint as:
- 3.9.1 the OSCE scoring framework is heavily weighted towards ill-defined subjective criteria ('domains') that expose the scoring process to unacceptably high risks of latent and patent racial bias (**OSCE domain complaint**);
 - 3.9.2 some OSCE confederates conduct the OSCE in a racially biased and discriminatory way which tends to disadvantage NCC (**OSCE confederate complaint**);
 - 3.9.3 OSCE examiners score candidates in a racially discriminatory way, and introduce their own perception bias and confirmation bias, in a way which tends to favour CC and disadvantages NCC (**OSCE examiner complaint**); and
 - 3.9.4 the College has a method of statistically modifying its results which amplifies the racially discriminatory nature of the raw results. It then fails to statistically modify these results to remove the element of clear racial bias from its results (**OSCE results complaint**).
- 3.10 In addition to the above discrimination allegations, the complainants alleged that the College denies candidates procedural fairness, through;
- 3.10.1 a failure to provide candidates with a transparent examination and grading process (**the transparency ground**);
 - 3.10.2 a failure to conduct the examination in a fair and impartial way that minimises the risks of bias and error (**the impartiality ground**); and
 - 3.10.3 a failure to provide reason and adequately account to candidates for their OSCE results (**the accountability ground**).

Principal Remedy Sought by Complainants

- 3.11 The complainants are seeking that the results of the 2016.2 OSCE be reviewed and statistically modified to remove any element of racial bias from the results. They suggested that various methods are available to detect and measure the impact of racial bias in the 2016.2 OSCE, specifically aimed at identifying OSCE stations where otherwise inexplicable statistically significant divergence between CCs and NCCs results (or the mean of those results) occurs.
- 3.12 The complainants contend that where a disparity in pass rates between NCCs and CCs has occurred, and the extent of the difference cannot be explained satisfactorily on the basis of validly assessed performance, and the variation is explicable by latent or patent racial discrimination, it becomes necessary to statistically modify the results at stations infected by discrimination to remove the discriminatory effect.

Psychometric Analysis of the 2016.2 OSCE Data

- 3.13 The EAG sought a psychometric analysis of the 2016.2 OSCE data to identify whether:
- 3.13.1 there was any statistical evidence of bias against Group B⁸ candidates;

8 Note that for the purposes of this analysis, Group A and Group B cohort references were used as this is the only demographic information available to the College by which to characterise the cohorts.

- 3.13.2 whether there is a propensity for specific examiners/stations to mark Group B candidates harder than Group A candidates; and
- 3.13.3 if this was occurring in particular domains more than others (such as communication, scholarship and teaching or leadership and management).
- 3.14 The analysis sought to evaluate the likelihood of the disparity in results being explicable by the following:
- 3.14.1 the groups are markedly different with respect to the ability the examination purports to assess; so the examination is valid and the difference that is highlighted is a true difference; or
- 3.14.2 the groups are not really different in ability and the examination is biased and therefore not valid; or
- 3.14.3 the result is a one-off effect which is due to a measurement error.
- 3.15 Results of this analysis undertaken by Professor Schuwirth indicate:
- 3.15.1 Chance/error is not the cause of the difference in pass rate between Group A (70.5%) and Group B (13.5%). Of the remaining two possibilities, there is no direct evidence to support bias against Group B candidates. Combining this with the assumption that, ‘during the station construction process these issues have been addressed as well and the stations have been scrutinised for any such possible bias’, Professor Schuwirth concluded that the difference observed is a true difference in performance.
- 3.15.2 Consistent differences in performance between Group A and Group B subsisted across all stations, domains and examiner background groupings; and no specific stations, domains, or examiners groups could be identified that would sufficiently and plausibly account for the difference in pass rates.
- 3.15.3 Although there is no way to determine whether any form of bias occurred in individual stations or by individual examiners (either against Group A or Group B), with the available data there is no indication of any form of bias big enough to statistically detect and account for the difference in pass rates.
- 3.16 Despite this the EAG is concerned about the marked disparity in results between Group A and Group B candidates that has grown since the introduction of the OSCE in 2015 and has had a significant impact on Group B candidates’ chances of successfully passing the Fellowship Examination.

Whether Unlawful Discrimination has Occurred

- 3.17 The EAG is not a court or statutory decision-making body, but rather an expert advisory group providing advice to the College. It is not the EAG’s role to make a finding or reach a formal determination in respect of whether the College has engaged in unlawful discrimination. However, in considering the complaint and advising on any remedies that should follow, the EAG has had regard to recent relevant court decisions on discrimination in fellowship examinations:
- 3.17.1 *The Queen on the application of Bapio Action Ltd v Royal College of General Practitioners and General Medical Council* [2014] EWHC 1416;
- In this case the complainants contended that the College and the GMC as the regulator, failed to fulfil the public sector equality duty placed upon them in that there was a “marked difference” in pass rates for the clinical skills assessment (final assessment) between doctors who obtained their medical degree in the UK and non-UK graduates or those that were black or from a minority ethnic group. While the Court accepted that requiring would be general practitioners to submit to the Clinical Skills Assessment put South Asians in both groups at a disadvantage, it found it was a proportionate means of achieving a legitimate aim principally because:
- (a) It was a high stakes assessment in that whoever passes will be licensed to practice medicine in an unsupervised capacity;
 - (b) The Clinical Skills Assessment is not a culturally neutral examination – it is designed to ensure Doctors are safe to practise in UK general practice;
 - (c) No better means of testing the requisite skills has been devised;

- (d) The method is in common use internationally; and
- (e) The method serves the legitimate purpose of protecting patient safety at a cost that is tolerable to those who ultimately successfully pass.

3.17.2 *Sklavos v Australasian College of Dermatologists* [2017] FCAFC 128;

The Full Court of the Federal Court of Australia affirmed that in considering the reasonableness of the relevant College's fellowship examination, the fact that there is an alternative assessment available that would accommodate the interests of the complainant and achieve the College's objective is a relevant consideration. However, it is not determinative of whether the College's chosen assessment method was unreasonable in the circumstances and ultimately discriminatory.

The EAG's Findings Relevant to Complaint

The domain complaint and the examiner complaint

- 3.18 The EAG considers that the introduction of the Fellowship OSCE in 2015 has had the unintended consequence of giving rise to systemic racial discrimination. Initially, the EAG thought this was attributable principally to unconscious bias of examiners, as outlined in the interim report to the College Board on 19 June 2017. However, the psychometric analysis of the 2016.2 OSCE data has found no indication of any form of examiner bias big enough to statistically detect and account for the marked disparity in pass rates between Group A and Group B candidates in the 2016.2 OSCE.
- 3.19 While the psychometric analysis has demonstrated that examiner bias did not have a significantly statistically identifiable effect on the 2016.2 OSCE, individual examiner bias could not be ruled out completely. The EAG has identified a number of factors that may have had an impact on the examination outcomes. As scoring of candidates performance involves independent expert judgement, it is potentially open to subjectivity and conscious or unconscious bias. The absence of documented marking criteria for candidate performance for 'minimum level of competence displayed' for marking Fellowship OSCE domains such as communication, professionalism, scholarship and teaching, and leadership and management can lead to greater individual subjectivity in the evaluation of these domains and interpretation of a candidate's performance as 'just at standard'. Others factors may also have played a part: relatively low examiner diversity (e.g. in the 2016.2 OSCE 93% of examiners were from Group A countries of primary medical qualification and 7% were from Group B countries of primary medical qualification), sub-optimal discussion of cultural sensitivity in calibration of OSCE assessment criteria, and the fact that raters' cognition is predominantly influenced by their own experiences, values and interests. Taken together, all these factors may have resulted in judgements giving rise to differential and potentially unfair outcomes for Group B candidates.
- 3.20 Following further enquiries since the interim report, the EAG has formed the view that examiner bias is not the principal reason for the disproportionate pass rates between Group A and Group B candidates. On the basis of evidence presented, clarification of the development of and changes to the OSCE process and observation of the examination process, the EAG has determined that the cause is multifactorial (these factors are set out in Section 7 - *Findings* in this report) with unconscious bias of examiners being one among many potential factors.

The confederate complaint

- 3.21 It was acknowledged in some submissions that actors and confederates are not immune to their own biases, which can create inconsistencies in the way in which they interact with candidates, impacting on candidate responses and performances in stations. This observation has been made by both trainees and examiners through the submission process.
- 3.22 The complainants and a few other submissions reported on their own experiences where they had felt that confederates had not been fair towards them or equitable in the interactions with all candidates.
- 3.23 Some of the submissions noted that the vast majority of confederates were from English speaking backgrounds and did not reflect the multi-racial presentations to many Emergency Departments.

- 3.24 Other than reports of individuals' own experiences, the EAG did not find any evidence of confederates conducting themselves in a racially biased way. A recent international study examining whether role players/simulated patients affect the outcomes of candidates' performances in high stakes OSCEs in terms of gender or ethnicity has found that the principal predictor of case score variance was the source of the candidates' primary medical qualification, and role player gender or ethnicity did not affect candidates' case scores.⁹ Significantly, the authors found non-marking role players are unlikely to be systematically influencing candidates' subgroup marks, however they did acknowledge that individual actors could conceivably be discriminating for or against any individual or group of candidates by making the case more or less difficult for them. The authors noted that in such situations the examiner would be expected to see this, and to mark and act accordingly.
- 3.25 The EAG accepts that individuals are not immune from their own biases and encourages the College to consider strategies for greater diversity in gender, age, ethnicity and language skills of the confederates and greater emphasis on training and monitoring of confederates/actors to ensure consistency and fairness in case presentation.

The results complaint

- 3.26 Advice received from experts¹⁰ was that the College's approach to standard setting and use of the borderline regression method is consistent with standard practice. Relevantly, Professor Farmer stated:

The borderline method is simple to execute in the time pressure of an OSCE. It is critical nonetheless that all examiners understand clearly the definition of a "borderline" candidate or in the case of the ACEM examination the meaning of the third point "just at standard", and are able to make a confident expert decision about whether a candidate falls below, at or above this category.

In order to make this judgement as robust as possible, and to ensure that examiners clearly understand the cognitive differences between a station score and a global rating, examiner training, including practice in using the marking method and calibration discussions, is essential.

- 3.27 Accordingly, the EAG does not consider that the College's use of this method amplifies the discriminatory outcomes within the results. Further, the EAG considers that statistical modification to remove racial bias in the results would only be appropriate where this is statistically detectable and significant.

Denial of procedural fairness

- 3.28 The EAG finds that since the introduction of the OSCE, unsuccessful candidates have not been provided with adequate feedback to enable them to understand the basis for their marks and identify areas for future improvement and practice.
- 3.29 Further, various concessions have been made by personnel within the College that the changes to the Fellowship Examination format and assessment could have been better communicated to trainees to enable them to better understand the process, how to prepare and how they would be assessed.
- 3.30 The College has acknowledged that training in cultural competence/sensitivity requires further focus both at trainee and FACEM level. Since the 2016.2 OSCE, the College has introduced a number of mechanisms that will contribute to minimising the risks of bias and error in examiner marking. They include:
- 3.30.1 the roll out of training modules on the topic of Assessing Cultural Competence, which has mandated that all teachers and assessors, including examiners, complete the modules;
 - 3.30.2 a written briefing about unconscious bias provided to all examiners;
 - 3.30.3 greater duration and emphasis in examiner workshops to clarify the marking criteria and determine a shared understanding of when a candidate's performance is 'just at standard'; and
 - 3.30.4 following 25% of candidates having rotated through stations, an additional calibration session is conducted to clarify and calibrate marking for consistency amongst examiners.

⁹ Denney ML, Wakeford R. Do role-players affect the outcome of a high-stakes postgraduate OSCE, in terms of candidate sex or ethnicity? Results from an analysis of the 52,702 anonymised case scores from one year of the MRCGP clinical skills assessment. *Education for Primary Care*, 2016; 27:1, pp.39-43.

¹⁰ Professor Elizabeth Farmer (22 August 2017) and Professor Lambert Schuwirth (September 2017).

- 3.31 These initiatives (and others introduced in 2017 to date) are a direct result of continuous monitoring, evaluation and improvement as part of the overall direction of the College and relate to all aspects of the OSCE, including examiner cohort, examiner training, candidate feedback and the use of technology, such as electronic marking, and the use of examination recording. The EAG considers that such continuous improvements are a positive start in addressing the complainants' concerns regarding transparency, impartiality and accountability of the Fellowship OSCE. The EAG makes additional recommendations later in this report also aimed at addressing these concerns.

Recommendations on Appropriate Remedies for the Complainants

- 3.32 The EAG invited the complainants to make further submissions on the remedies they were seeking. Through this process the complainants offered up a number of suggestions aimed at reducing susceptibility of bias in the OSCE:

Remedies requested by complainants

- 3.33 The 2016.2 OSCE be reviewed and statistically modified to remove any element of racial bias from the results;
- 3.34 Marking criteria for individual domains tested at each station and their relative weighting should be detailed, so as to guide examiners on the relevant variables to mark candidates;
- 3.35 Multiple examiners should be used at each station to allow for a diversity of views and opinions in marking;
- 3.36 Audio-visual recording of examinations should be conducted routinely for review and checking for inter-rater reliability;
- 3.37 Use of independent observers or scrutineers to review station designs, and attend and observe OSCEs;
- 3.38 An internal review of marking should be conducted as a matter of course, where a failing candidate is within one standard error measurement;
- 3.39 Review of examiners' mark sheets for reliability between candidates and internal consistency;
- 3.40 Ongoing cultural awareness education for examiners;
- 3.41 Encourage and support diversity in appointments to the Court of Examiners and other senior positions in the College;
- 3.42 Monitoring of examiners' performance to detect racially discriminatory tendencies in scoring;
- 3.43 Reciprocal feedback and assessment of examiners and confederate conduct by candidates; and
- 3.44 Candidates receive a random assigned identification number to identify them during the examination.

EAG's response to remedies requested by complainants

- 3.45 Given the psychometric analysis of the 2016.2 OSCE data has found no indication of any form of examiner bias statistically significant enough to detect and account for the marked disparity in pass rates between Group A and Group B candidates in the 2016.2 OSCE, the EAG has determined it is not necessary or appropriate to make statistical adjustments to the 2016.2 OSCE results.
- 3.46 However, the EAG was greatly assisted by the other suggestions made by the complainants, which have influenced the formulation of the EAG's recommendations relating to the future conduct of the College's examinations and related College processes.

EAG's recommended remedies for the complainants

- 3.47 The EAG acknowledges the impact the disproportionate pass results have had upon unsuccessful candidates and recommends the following remedies for the complainants:
- 3.47.1 An apology from the College for the unintended systemic racial discrimination associated with introduction of the Fellowship OSCE in 2015 and for the inadequate feedback they received on their OSCE results.
 - 3.47.2 An offer from the College to refund examination fees for the 2016.2 OSCE to the affected complainants.
 - 3.47.3 To aid those contemplating a further attempt at the OSCE:
 - (a) extend the time for completion of training, particularly for those who are nearly time expired;
 - (b) provide tailored/structured formative feedback on past examination attempts, to the extent that it is able to;
 - (c) offer a College run/endorsed OSCE preparation training, including examination psychology preparation;
 - (d) consider videotaping subsequent attempts and providing feedback using video footage; and
 - (e) allow extension of time for formal requests for review and/or appeal from complainants, notwithstanding that time limits may have expired.
 - 3.47.4 To aid those contemplating leaving or who have left the training program:
 - (a) provide career counselling; and
 - (b) offer the pathway to qualification for Diploma in Emergency Medicine through recognition of prior learning pursuant to the applicable College policy.
- 3.48 The anonymity of the complainant group and imprecise characterisation of the affected cohort has made recommending suitable remedies for the complainants challenging. Accessing the above remedies will inevitably require the complainants to identify themselves to the College as having been affected by the 2016.2 OSCE. The EAG expects the College to facilitate access to these remedies confidentially and that the individuals who choose to access the remedies will not be subject to any resulting discrimination or victimisation.
- 3.49 Although the complainants are a smaller group, they are representative of trainees impacted by the College's transition to the Fellowship OSCE. Accordingly, the EAG suggests that the College consider extending the remedies proposed for the complainants to all affected candidates from the 2016.2 OSCE.

4 Submissions

- 4.1 The EAG has sought to understand individuals' experiences with the College examination processes, with a key focus on the OSCE.
- 4.2 Facilitated through an advertisement for submissions, oral hearings before the EAG and oral hearings with independent members of the EAG, a comprehensive set of submissions has been received covering College examination processes, ITAs and WBAs, with a particular focus on the OSCE.
- 4.3 The EAG has received more than 60 written submissions from trainees, examination candidates, FACEMs, DENTs and examiners. The independent members of the EAG have also received a further 26 direct written submissions from trainees, FACEMs and DENTs about their experiences with the Fellowship Examinations.
- 4.4 Submissions were also received following the limited publication of the interim report. It was at this time that a number of current examiners made submissions on the preliminary findings in the interim report, including in some cases indicating agreement with those preliminary findings. Examiners also participated in an oral hearing where they discussed the issues as they saw them, including matters which had arisen from the interim report.
- 4.5 The key themes in the submissions were the subjectivity of the OSCE and scoring of the OSCE, inadequacy of preparation and training, inadequate feedback, the overall conduct of the examiners, confederates and examiners, reasons for differential results, prevalence of stereotypes and the toll on trainees.
- 4.6 Of note is that submissions received from trainees, examination candidates, DENTs, FACEMs and examiners, broadly focused on the OSCE process as a whole, rather than simply the approach to the 2016.2 OSCE. Due to the nature of submissions received, it was not possible to differentiate between submissions focused solely on the 2016.2 OSCE and those focused more broadly.

Submissions from Trainees and Examination Candidates

- 4.7 Submissions were received from current trainees and examination candidates who demonstrated that they had been significantly impacted by the current format of the OSCE, both professionally and personally. In some cases, examination candidates had failed more than three times, with more than one examination candidate stating that they would exit the training program due to the impact their repeated failure was having on their lives. Candidates also stated that they had failed by 1% or 2% on more than one occasion. The submissions received have been analysed and key themes recurrent throughout the submissions have been identified. These key themes are set out below and a discussion on each theme is included.

Subjectivity and scoring

- 4.7.1 The apparent subjectivity of scoring in the OSCE has led to the perception that NCCs are in some way negatively affected by the OSCE process, be it intentional or unintentional. The processes for marking and scoring the OSCE are subjective as examiners are required to assess whether candidates meet broad objectives of the individual station, without an express set of criteria to mark candidates against.
- 4.7.2 To some this process is viewed as examiners assessing candidates by "feel" rather than a quantitative method. This is compounded by the use of the borderline regression method used to standardise results, and a lack of a known pass mark for the OSCE until after it has occurred.
- 4.7.3 There is also a perception that the OSCE process was introduced as a method to remove NCCs from the Fellowship program.

Adequacy of preparation and training

- 4.7.4 There is a perception that examination candidates are inadequately prepared for the OSCE, which adversely impacts their outcome. This perception arises in light of the limited resources produced by the College to assist with OSCE preparation, including that candidates are not provided with a station breakdown, sample examination scenarios or the pass mark prior to undertaking the examination.

- 4.7.5 Additionally, there is also a perception that the ITAs and WBAs undertaken throughout training ought to adequately prepare candidates to take the Fellowship Written Examination and OSCE.
- 4.7.6 Instead, trainees believe they are not provided with adequate feedback during their training to identify areas for improvement, including those areas which could ultimately hold them back during Fellowship examinations.
- 4.7.7 Following the introduction of the OSCE and the limited information provided by the College, candidates submit they have undertaken a multitude of examination preparation courses, including informal study groups, formal groups at training hospitals and courses run by the College. It however remains a perception among the Group B candidates, that trainees are inadequately prepared to undertake the OSCE in the current format, despite believing their preparations were adequate.
- 4.7.8 Such perceptions have been affirmed by DEMENTs and FACEMs who have been unable to provide examination candidates with specific guidance on how they ought to prepare themselves for the OSCE; in some cases this is due to a candidate failing the OSCE whom the DEMENT believed was at the required standard to succeed.

Inadequacy of feedback

- 4.7.9 Trainees and candidates submit that the lack of feedback after undertaking an OSCE substantially impacts on their future preparedness to retake the examination. This continues to exacerbate the perception that the OSCE is being used to limit the candidates who pass the Fellowship Examinations and become FACEMs.
- 4.7.10 Where feedback is given, candidates submit that it varies greatly between examiners, is at times defensive rather than constructive, and more often than not addresses personality traits of candidates, rather than clinical competencies.
- 4.7.11 As a result, candidates have sought additional feedback and support from their DEMENTs and/or other FACEMs they work with. DEMENTs and FACEMs have submitted in support of examination candidates, that they do not receive additional feedback for their trainees from the College, and have been told that they are unable to obtain feedback even where expressly requested.

Conduct of examiners, confederates and actors

- 4.7.12 Trainees and candidates also expressed concerns regarding the conduct of examiners, confederates and actors in various OSCE stations.
- 4.7.13 Concerns included that single examiners did not have auditable results and that single examiner stations allow subjective biases to infect conduct.
- 4.7.14 Confederates and actors were reported to have treated CCs and NCCs differently when presenting scenarios. For example, one actor reportedly pretended they could not understand the candidate, who had a slight Asian accent. Further, one actor was reported to have told different candidates different histories, and ultimately the NCC was disadvantaged by this action.

Prevalence of stereotypes

- 4.7.15 Trainees and examination candidates have also made submissions surrounding the belief that some DEMENTs, FACEMs and examiners have regarding certain stereotypes of NCCs.
- 4.7.16 These stereotypes have been submitted to impact a trainee's assessment in that examiners automatically believe that a certain stereotype, such as being quietly spoken, means that the examination candidate is not confident about their decision. Such stereotypes ultimately result in a belief that the candidate is not suitable for Fellowship, despite being otherwise capable within the hospitals they work, of undertaking and performing to a high standard and commonly without supervision.

Personal toll

- 4.7.17 There has been a significant personal toll on candidates with many experiencing stress and depression. Experiences include significant financial impact in that they are required to pay for each time they undertake the examination, to take time off work to study for the examination and, in some circumstances where the examinations take place interstate from their home base, to pay for flights as well as travel expenses.

Submissions from FACEMs and DENTs

- 4.8 Submissions were received from approximately 20 FACEMs and DENTs. Submissions received included FACEMs making submissions on their own experiences in the recent examination processes. The submissions received were analysed and collated into key themes recurrent in the submissions. As set out above, below are the key themes recurrent in submissions received from FACEMs and DENTs.

Adequacy of preparation and training

- 4.8.1 The general perception among FACEMs and DENTs is that limited guidance on the process, content and marking for the OSCE being made available to the candidates prior to the OSCE impacts on their preparedness.
- 4.8.2 Submissions received indicate there is a perception among DENTs that candidates believe that if they are appointed as an ED Registrar, they are capable of working as a FACEM and therefore automatically able to pass the Fellowship examinations.
- 4.8.3 DENTs are concerned that trainees do not understand that FACEMs need a higher order of clinical, communication and administrative skills to be able to manage life and death decisions rapidly that ED Registrars may not be equipped with.

Inadequacy of feedback in training

- 4.8.4 Further, FACEMs and DENTs fail to give honest feedback about a candidate's clinical skills during training. This issue embraces a number of points, including that:
- (a) DENTs and FACEMs are not giving feedback early enough in training, including feedback that trainees do not have the skills or knowledge to continue in the training program or are not yet prepared to take the Fellowship examinations; and
 - (b) where a trainee has progressed to the point of undertaking the Fellowship examinations, DENTs and FACEMs perceive that the examinations themselves will appropriately pass or fail those candidates who are not capable of progressing to Fellowship.
- 4.8.5 DENTs have also acknowledged that some trainees are more receptive to feedback than others. In one reported instance, a trainee was not receptive to a female DENT's feedback on their performance.
- 4.8.6 FACEMs and DENTs also submitted that while formal feedback may not have been given in response to ITAs and WBAs, informal feedback had been given (at least to some trainees) and this might not have been adequately recorded.

Inadequacy of feedback in Fellowship Examinations

- 4.8.7 Candidates' overall preparedness is frustrated by a lack of specific feedback for those who fail one or more OSCE. A lack of feedback means that FACEMs and DENTs are unable to assist their trainees to adequately prepare to retake the examinations.
- 4.8.8 Trainees and DENTs have found it extremely difficult to obtain feedback from the College following Fellowship Examinations. A specific example was given when a number of trainees failed a station where they were required to read an ECG. The trainees received no feedback and asked for assistance from their DENT. The DENT sought feedback on their behalf and was told that the information was confidential and could not be disclosed. Another FACEM sought feedback for a specific trainee and was told that the trainee's performance "might have been good enough in India but was not good enough here". Such experiences are not limited to these examples.

Reasons for differential results

- 4.8.9 FACEMs submitted that clinical competency or lack thereof of the trainees attempting OSCEs was a potential rationale for the low pass rates of NCCs. There is a perception that poor performing trainees who have consistently demonstrated issues with their clinical competencies have been inadequately managed throughout their training.
- 4.8.10 FACEMs have experienced that some trainees are not at the standard required to be admitted to Fellowship, and raised concerns about the processes permitting such candidates to undertake the examinations.
- 4.8.11 It was also submitted that the very nature of an examination is to discriminate between those who can and those who cannot pass it based on their clinical competencies, rather than some other characteristic candidates may share. Accordingly, the clinical examinations will by its very nature “weed out” those who are not capable of practising as FACEMs.¹¹
- 4.8.12 One DEMENT noted that, prior to the introduction of the new examination processes, NCCs struggled with the written examination, which at the time tested complex reasoning and management through essay style questions. The same DEMENT opined that the change reduced the possibility of bias and dependence on English language proficiency thereby making the written examination easier for NCCs to pass.
- 4.8.13 Further, it is more common to identify inherent issues in trainee performance in the later stages of training when the level of experience has evened out between all trainees and issues become more apparent.
- 4.8.14 Candidates who have English as a second language are automatically disadvantaged by the OSCE process given short time frames to respond to scenarios as well as different communication styles, interpretation of social cues and cultural norms.
- 4.8.15 Nevertheless, it was submitted that a core skill of a FACEM is to be capable of taking a detailed and accurate history from a patient, assessing the patient and communicating the nature of the patient’s presenting problem. Additionally, the role includes communicating with departmental staff about patient care. If a trainee is not capable of meeting these fundamental requirements of an emergency physician working in Australia or New Zealand as applicable, they should be failed as it would be negligent to do otherwise.

Submissions from Examiners

- 4.9 Following the interim report, a number of examiners made submissions on the terms of reference as well as in response to the interim report. Submissions were made in writing as well as in oral hearings before the Expert Advisory Group.

Concerns regarding the interim report/interim findings

- 4.9.1 The examiners were concerned about the preliminary conclusion that unconscious bias among examiners is the primary mechanism for the effects experienced by NCCs. The examiners were surprised given examiner orientation, training and performance monitoring that occurs as a standard process. It was also opined that if unconscious bias is present, it would have to be widespread among the examiners to have such an impact given that each examiner accounts for a very small part of a candidate’s overall mark.
- 4.9.2 Examiners were also concerned about the matters raised in the interim report about examiner selection criteria. The selection criteria necessarily require prospective examiners to demonstrate their capacity to examine trainees. The pool of prospective examiners is limited to those Fellows who already participate in College activities. Of note is one examiner’s concerns that if a requirement of cultural diversity within the Court of Examiners was imposed, advice would be required on how this requirement can be implemented without introducing another form of discrimination of one cultural group being preferred over another.

11 Of note is that many trainees and examination candidates submitted that while working as trainees in emergency departments, they were often left to run the department without FACEM supervision overnight. These submissions suggested there was no cause for concern about trainees operating the department without supervision.

Examination process

- 4.9.3 As stated above at 4.9.1, submissions highlighted the existing processes in place to limit/remove any biases that could affect the outcome of an examination. Examiners believe these processes include a marking system that requires candidates to satisfy specific criteria, thereby removing bias.
- 4.9.4 Submissions indicated that the Fellowship examinations as a whole have become a much higher stakes process than in previous years as the availability of jobs for Fellows has decreased.

Adequacy of preparation and training

- 4.9.5 The process by which the OSCE was introduced was flawed as it was poorly communicated and candidates were provided with little warning of the change. This has been further complicated by wide variations in cut scores, an inaccurate method of assessing specialist emergency medicine practice knowledge, very high pass rates for the 2015.1 Written Examination across all candidate pools, removal of feedback for unsuccessful candidates and removal of specific information about station content. The new style of examination also saw a change in the weighting of specific domains over others, so for instance in the 2015.1 OSCE, communication accounted for 30% of a candidate's overall mark as it was tested in each station. However, in previous examinations it had only accounted for 5% of a candidate's overall mark. In one examiner's submission, this more than any other factor has resulted in a marked change in pass rates for Group B candidates.
- 4.9.6 In addition to the changes in the oral examination, examiners have pointed to substantial changes to the format of the Written Examination. These changes have resulted in a less analytical assessment of deeper knowledge than was found in the previous written examination. The data has demonstrated that the percentage of Group B candidates passing the Written Examination has remained largely unchanged despite the introduction of the OSCE, with the exception of the Group B cohort who sat the 2015.1 Fellowship Written Examination where an 80% pass rate was achieved. Those candidates who would have failed that part of the examination in the past are progressing to the oral examination and then failing.
- 4.9.7 Examiners submit that the change to an OSCE meant that the focus changed to a socio-linguistic performance test with no ability for the examiners to explore issues arising during the stations, due to limited if any examiner-to-candidate interaction. The change from an examiner led question and answer style of oral examination to a minimally interactive examiner style examination had a significant impact. This has also affected an examiner's ability to assist candidates to demonstrate their actual knowledge.
- 4.9.8 It is broadly submitted that issues related to the examination process and change in examination style greatly impacted examination candidates, and, in particular, the ability of Group B candidates to pass.

Reasons for differential results

- 4.9.9 Examiners submitted that while there has been a marked decrease in the pass rates of Group B candidates since the introduction of the OSCE, there has been no change in the composition of the Court of Examiners, methods of examiner training or the examiners' code of conduct over this time.
- 4.9.10 Examiners submit that the issues experienced by Group B candidates were compounded by the matters set out above in 4.9.5 in that with all these issues, there were further changes decreasing the already limited interaction between examiners and candidates, leading to the perception that there must be bias causing the low pass rates. Additionally, with these changes, candidates lost faith in the College process and believed that the College's motivation for the examination process had changed and that it was an attempt to remove trainees from the program rather than assess clinical competency to practice.

- 4.9.11 Examiners also noted that the issues experienced by Group B candidates can be examined more broadly, based on where a candidate's primary medical degree was obtained. For countries/ regions with an OSCE style examination (including parts of the United States), those candidates will be more familiar with that process from their primary medical degree. Conversely, candidates who have not been educated using an OSCE process (including parts of India and Pakistan) will find it more difficult to achieve a pass in the OSCE.
- 4.9.12 Initially, examiners have acknowledged that Group B candidates (and Group A candidates who had failed) were significantly disadvantaged by a lack of useful and constructive feedback about the issues identified with their clinical competency. Feedback provided only addressed marks for stations that accounted for a certain percentage of their overall mark. Since the interim report, changes have been introduced to provide candidates with feedback about what they have failed to address in the OSCE.
- 4.9.13 Examiners also note that pressures of job maintenance sway candidates to take examinations when they are not necessarily ready. While DEMENTs can recommend to trainees that they should not take the examinations for Fellowship yet, there is no requirement that candidates obtain their DEMENT's approval prior to taking the examinations. In many situations, trainees are not adequately prepared or experienced, which increases the chances of failure.

5 Analysis of Differential Pass Rates

- 5.1 The complaint alleged that only 6.8% of NCCs passed the 2016.2 OSCE compared with 88% of CCs. Data provided to the EAG (refer Tables 4 - 6 and Figure 2, pp.34 - 36) and produced in the College's Reaccreditation Submission to the Australian Medical Council and the Medical Council of New Zealand indicates that from the time of the introduction of these changes to the Fellowship examination format in 2015, a marked disparity in the pass rates between Group A and Group B candidates has occurred.¹² Up until 2015, the data shows some disparity in the pass rates between these groups, however, the disparity in pass rates became markedly greater following the introduction of the new examination format (refer Table 5, p.34).
- 5.2 It is evident that the change in the Fellowship Clinical Examination process and the introduction of the OSCE in 2015 has had a disproportionate impact on Group B candidates, reflected in their considerably lower pass rate.
- 5.3 Some submissions believe this is due to racial biases. Most submissions consider bias unintentional; however, some believe that it is intentional. The EAG has found no evidence that the introduction of the new OSCE format was to intentionally disadvantage a particular cohort of trainees. The EAG has formed the view that many factors are likely to have contributed to the differential pass rate of this cohort – as discussed below.

A Review of Literature on Success Rates and Examiner Bias in High-Stakes Clinical Examinations

- 5.4 A literature review conducted by Professor Elizabeth Farmer for the EAG (referred to herein as the "Farmer review") demonstrates that the College is not alone in its experience, with various international studies revealing a persisting difference in attainment between local and international medical graduates in respect of high-stakes examinations ranging from medical school to high-stakes postgraduate specialty examinations. Professor Farmer's review illustrates that these differences have been reported consistently and repeatedly over several decades in multiple countries and in multiple disciplines with similar patterns including Australia.
- 5.5 The Farmer review found that postgraduate specialty examinations reported in worldwide literature indicate that graduates who are citizens of that country and trained in that country perform better than graduates trained in other countries. Somewhat concerning however is that this difference persists even among graduates who are citizens of that country but belong to different ethnic minority groups, for example British studies have shown that white UK graduates perform better than South Asian or Black Minority Ethnic (BME) UK graduates.
- 5.6 The Farmer review noted that very few studies have investigated the potential for examiner bias specifically arising from the perceived country of training or ethnicity in high stakes examinations in relation to lower success rates of IMGs in comparison to local graduates. Most studies were quantitative with a focus on testing for bias using actual examination data in large-scale data sets.
- 5.7 The literature discussed in the Farmer review highlights that findings of differential attainment cannot be dismissed as atypical or local to any one country or specialty examination. However, the College's experience of such a marked disparity in results following the introduction of the OSCE examination format in 2015 does appear anomalous compared with other reported experiences.

¹² ACEM Reaccreditation Submission to the Australian Medical Council and the Medical Council of New Zealand, June 2017, p.131.

- 5.8 The following is extracted from the Farmer review:
- 5.8.1 In Britain, a difference between candidates' success in undergraduate and postgraduate high-stakes examinations by ethnic background and gender has been reported over the past two decades.¹³
- 5.8.2 Esmail and Roberts have investigated the difference in failure rates in the postgraduate examinations of the Royal College of General Practitioners (MRCGP qualification to practise independently in general practice) UK by ethnic or national background, and attempted to identify any factors that might be associated with these differences in the clinical skills (OSCE) component of the examinations¹⁴. The authors reported that, after controlling for age, gender, and level of performance in the written knowledge test, significant differences persisted between white UK graduates and other candidate groups. Black and minority ethnic graduates (BME) even though they were trained in the UK were significantly more likely to fail the clinical skills assessment at their first attempt than their white UK colleagues (failure rate 17% v 4.5%). Black and minority ethnic candidates who trained abroad were also more likely to fail the clinical skills assessment than white UK candidates (65% v 4.5%). The authors postulated that "subjective bias due to racial discrimination" in the clinical skills assessment may be one cause of failure for UK trained candidates and international medical graduates. They also postulated that examination success could be affected by extraneous factors, for example, in training experience and in other cultural factors affecting some candidates trained in the UK and abroad. They recommended that consideration be given to strengthening postgraduate training for international medical graduates.
- 5.8.3 Wakeford et al published a major British study concerning the difference in performance in specialty examinations for two different colleges (physicians and general practitioners). The focus of the study was the difference between white candidates and black minority ethnic (BME) candidates. The main difference in this study compared to others is that both knowledge tests and clinical tests were examined, and these also were compared for two entirely independent testing organisations.
- 5.8.4 The detailed analyses by candidate ethnicity show that although white candidates outperformed BME candidates, the differences were largely mirrored across the two different sets of examinations. Although the reason for the differential performance is unclear, the authors concluded that:
- "... the similarity of the effects in independent knowledge and clinical examinations in different specialty colleges suggests the differences are unlikely to result from specific features of either assessment and most likely represent true differences in ability."* (p.1).
- 5.8.5 This study found a negative ethnicity effect at each stage of both independent examinations; BME candidates performing less well even after taking performance at previous stages into account. The authors considered that these effects are therefore unlikely to be due to particular features of any one assessment, component of an assessment or style of assessment. The finding that similar effects are found on written tests suggest that these effects cannot be explained simply by bias on the part of clinical examiners.

13 Footnotes 26 - 33:

26 Tyrer SP, Leung W-C, Smalls J, Katona C. The relationship between medical school of training, age, gender and success in the MRCPsych examinations. *Psychiatr Bulletin* 2002;26:257-63.

27 Hurst NG, McManus IC, Mollon J, Dacre JE, Vale JA. Performance in the MRCP(UK) Examination 2003-4: Analysis of pass rates of UK graduates in the Clinical Examination in relation to self-reported ethnicity and gender. *BMC Medicine* 2007;5:8.

28 Wakeford R. International medical graduates' relative under-performance in the MRCGP AKT and CSA examinations. *Education for Primary Care* 2012;23:148-52.

29 Wakeford R, Farooqi A, Rashid A, Southgate L. Does the MRCGP examination discriminate against Asian doctors? *The BMJ* 1992;305:92-4.

30 Esmail A. Ethnicity and academic performance in the UK. *The BMJ* 2011;342:d709

31 Woolf K, et al. Exploring the underperformance of male and minority ethnic medical students in first year clinical examinations. *Advances in Health Sciences Education* 2008;13(5):607-616.

32 Schleicher I, et al. Examiner Effect on the Objective Structured Clinical Exam - A Study at Five Medical Schools. *BMC Med Educ* 2017;17(1):71.

33 Woolf K, Potts HWW, McManus IC. The relationship between ethnicity and academic performance in UK-trained doctors and medical students: a systematic review and meta-analysis. *The BMJ* 2011;342.

14 Esmail A, Roberts C. Academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between November 2010 and November 2012: an examination of data. *The BMJ*. 2013;347.

- 5.8.6 Richens et al have studied the UK and Ireland intercollegiate specialty board fellowship examinations, developed and conducted in the UK and Ireland by the intercollegiate boards of the Royal College of Surgeons UK and Ireland. They explored effects of gender, ethnic origin, first language, and training status on scores in these examinations across the computer-marked written section and in the face-to-face oral and clinical section. Demographic characteristics and examination results from 9,987 attempts across 177 sittings from 2009 to 2013 were analysed in an analysis of variance by training status, gender, ethnic origin, first language, and section (computer-marked multiple-choice examination vs face-to-face oral and clinical examination).
- 5.8.7 The authors concluded that training status was the most important factor in candidates' results. Although the analysis showed significant effects of ethnic origin and first language within 'core candidates', these differences were statistically indistinguishable between the two sections of the examination, suggesting that the differential attainment by these factors cannot be attributed to examiner bias in a face-to-face examination.
- 5.8.8 Jasper et al have reported on the different pass rates of the Royal Australian College of General Practitioners (RACGP) Fellowship examination from 1999 to 2004, for various candidate routes to the examination including local graduates and IMGs who have undertaken prescribed general practice postgraduate training programs across Australia.
- 5.8.9 While the figures and any tests of significance were not reported, the table (on page 28 of the Farmer review) clearly demonstrates that there was a substantial difference in the pass rates between local graduates and IMGs despite having undertaken the same postgraduate training programs. Similarly, the pass rates of IMGs who had not undertaken the training program (practice eligible route) were lower than Australian graduates who had not undertaken the training program.
- 5.8.10 Karnik et al have recently published an Australian study relevant to this review that attempted to distinguish international medical graduates by broad categories of country of training. They analysed the performance and predictors of success in the final fellowship examination of the College of Intensive Care Medicine (CICM), and specifically compared the outcomes for international medical graduates attempting the CICM fellowship examination with those of local trainees, defined as those from Australia, New Zealand and Hong Kong (ANZ-HK). They also compared the performance of IMGs from countries with comparable health care systems (CHS) with those from other countries (non-CHS).
- 5.8.11 In total, 233 candidates presented to the examination 334 times, and most (73%) were IMGs. ANZ-HK trainees performed significantly better at the examination (79% v 46%, $P < 0.0001$). IMG trainees from CHS performed significantly better than trainees from non-CHS (60% v 40%, $P < 0.01$). Any candidate completing an ANZ primary examination performed significantly better than non-ANZ primary candidates (74% v 41%, $P < 0.0001$). IMG candidates successful at a postgraduate examination from a CHS country performed significantly better than candidates from a non-CHS country (56% v 34%, $P = 0.005$). The authors concluded that a significant proportion of candidates appearing for the CICM fellowship examination are IMGs. There were differential pass rates of IMGs according to whether their country of graduation was from a comparable health system or not.
- 5.8.12 In addition, pass rates for trainees who graduated from the ANZ-HK systems had a higher success rate in the fellowship examination. IMGs from a CHS country, or those who completed an ANZ primary also had a much higher success rate compared with other IMGs.
- 5.9 The Farmer review observed that the causes of differential attainment remain unclear and while some effects of examiner and candidate interaction have been determined, they appear relatively small in comparison to the size of the disparity in results between groups of local graduates and international graduates.

The College's Perspective on the Variability in Pass Rates

- 5.10 For the purposes of compliance with AMC accreditation the College is required to:
- 5.10.1 Regularly review the quality, consistency and fairness of assessment methods, their educational impact and their feasibility (Accreditation Standard 5.4.1); and
 - 5.10.2 Use a recognised standard setting method to determine the pass mark of examinations such as the Fellowship Examination (Accreditation Standard 5.2.3).
- 5.11 The College has publicly acknowledged that the pass results clearly indicate significant differences between the pass rates of candidates from country Groups A and B on both the Fellowship written and clinical examinations over a period of time; which was the case generally for the examinations in the format used prior to 2015, as well as from 2015.¹⁵
- 5.12 The College has acknowledged that the data indicates:¹⁶
- 5.12.1 pass rates for the Clinical component of the Fellowship Examination since the implementation of the revised format are lower than desired; and
 - 5.12.2 a clear disparity exists between pass rates for the two groups over time, a trend that had not been properly identified until the monitoring of pass rates became more systematic following the implementation of the new format examinations.
- 5.13 The EAG is advised that pass rates (along with other aspects of College examinations) are now monitored more intensively and systematically than in the past, in an attempt to understand the possible causes of the trend in disparity of pass rates, as well as any other identifiable issues.
- 5.14 Data analysis undertaken by the College after each Fellowship examination includes the study of examination psychometrics, such as reliability, and candidate information, such as number of past examination attempts.
- 5.15 The College has advised the EAG that its analysis suggests candidate factors to be significant in the low pass rates associated with the Fellowship examinations, such as:
- 5.15.1 *Previous attempts at the OSCE* – The College's experience is that the vast majority of candidates who pass the Fellowship OSCE usually do so within three or four attempts. At the same time the more attempts a candidate undertakes, the less likely they are to pass. The College's data demonstrates that the proportion of candidates who sit the OSCE, who have had previous attempts at the clinical examination, has been steadily increasing over time. Further, the College considers that the commencement of the new Regulations from 2018, that limits examination attempts to three, may also have triggered more 'speculative' attempts at the OSCE by under-prepared candidates in order to gain the benefit of additional unlimited attempts.
 - 5.15.2 *Country of Primary Medical Degree* – The College's experience (and that demonstrated in other studies) is that there is an association between the country of origin in which a trainee undertook their primary medical degree, and the challenges in passing the OSCE. The College's data represented in Figure 1 (overleaf) shows the OSCE clear pass group size (%), against the proportion of trainees in the Group B category sitting each clinical examination. Analysis undertaken by the Colleges shows a strong negative correlation (Pearson correlation -0.610), leading to the interpretation that where a larger Group B cohort attempts an OSCE, there is a strong correlation with a smaller proportion of trainees overall ending up in the clear pass group. Significantly, the College's data on comparative pass rates in the written examination demonstrates a consistent disparity in pass rates between Group A and Group B candidates in that examination type (refer Table 4, p.34).

¹⁵ ACEM Reaccreditation Submission to the Australian Medical Council and the Medical Council of New Zealand, June 2017, p.123.

¹⁶ ACEM Reaccreditation Submission to the Australian Medical Council and the Medical Council of New Zealand, June 2017, pp.130-131.

5.15.3 *Other examination measures* – The College’s analysis of data after 2015, of candidates in the borderline group,¹⁷ found a strong negative correlation between the overall OSCE pass rate and the borderline group size (Pearson correlation -0.52) leading to the interpretation that a larger borderline group is strongly correlated with a lower pass rate. Further studies of the borderline group size found a strong negative correlation between the size of the borderline group and examination reliability (Pearson correlation -0.54) and a moderate negative correlation between the borderline group size and station reliability (Pearson correlation -0.44). Meaning that if there is a large borderline group, it is strongly associated with a lower overall examination pass rate, strongly associated with poor examination reliability and moderately associated with poor station reliability. A statistically significant negative correlation was found between the borderline group size and station discrimination¹⁸ (Pearson correlation -0.84, $p < 0.01$), which suggests that a larger borderline group is strongly correlated with a reduced ability of stations to discriminate between clear passing and clear failing candidates. It may also be possible that the method used to measure station discrimination becomes less accurate with examinations with larger borderline groups.

5.16 The following figure, Figure 1, illustrates the variability in OSCE pass rates between each examination conducted since 2015 for all trainees sitting and who obtained clear passes, and trainees in Group B.

Figure 1 *Variability in OSCE pass rates between each examination conducted since 2015 for all trainees sitting and who obtained clear passes, and trainees in Group B*

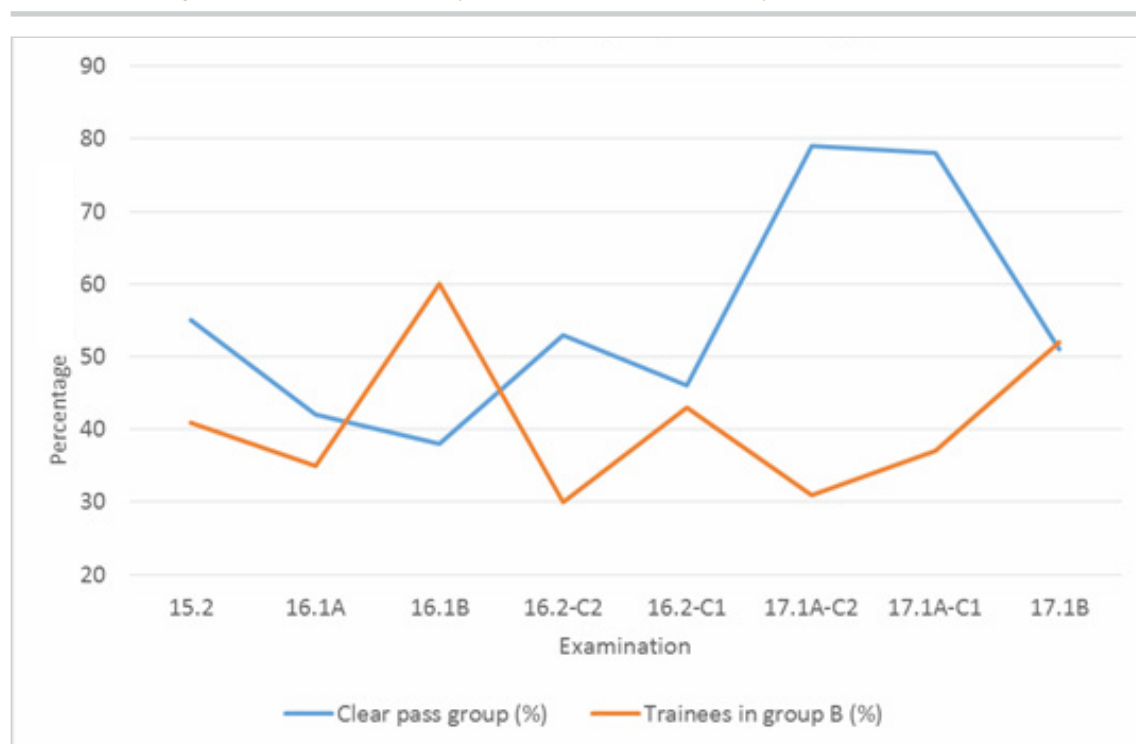


Figure 1 (above): Percentage of trainees sitting with results in the clear pass group and percentage of trainees sitting in the country ‘Group B’ category, which refers to trainees who obtained their primary medical degree from countries other than Australia, Canada, Ireland, New Zealand, UK and USA. C1 refers to cohort 1, C2 to cohort 2. Data from 2015.1 was not included, as SEM was not used in determining results in the examination.

17 Candidates who are considered a fail in the examination, since they have not clearly demonstrated that they have reached the required standard. This group includes candidates with final scores falling in the range of the raw cut score +/- SEM.

18 Discrimination in this context is a measure of how well a station is able to distinguish between examinees who are knowledgeable and those who are not.

Table 4 Fellowship Clinical Examination - Percentage of candidates by country group

	Total trainees in clear pass group (%)	Total trainees in country Group A (%)	Total trainees in country Group B (%)
2015.2	55	59	41
2016.1A	38	40	60
2016.1B	42	65	35
2016.2-C1*	46	57	43
2016.2-C2*	53	70	30
2017.1A-C1*	78	63	37
2017.1A-C2*	79	69	31
2017.1B	51	48	52

* C1 refers to cohort 1; C2 to cohort 2

Table 5 Fellowship Clinical Examination - Historical measures

	Pass rate (%)	Final cut score (%)	Mean score (%)
2015.1-C1*	76.9	61.7 [§]	65.9
2015.1-C2*	70.7	61.0 [§]	65.7
2015.2	55.6	63.6	63.4
2016.1A	42.4	64	62.1
2016.1B	38.0	63	57.1
2016.2-C1*	46.2	63	61.7
2016.2-C2*	53.1	64	63.9
2017.1A-C1*	77.6	64	70.7
2017.1A-C2*	79.1	64	68.3
2017.1B	50.7	63	62.7

* C1 refers to cohort 1; C2 to cohort 2

[§] SEM was not added to the cut score in these examinations.

Figure 2 Pass rate trends by country group per clinical examination

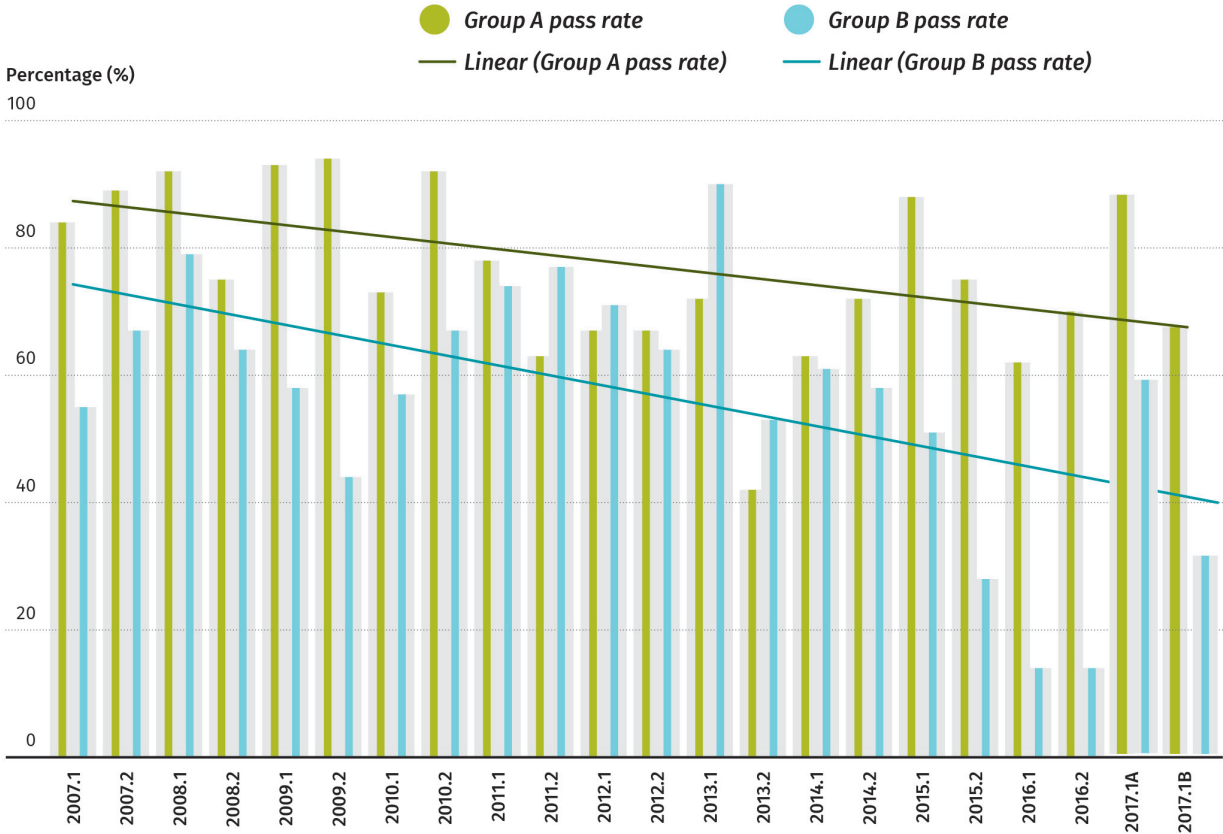


Table 6 Pass rates by country group per clinical examination

Examination	Group A				Group B			
	Total	Fail	Pass	Pass Rate	Total	Fail	Pass	Pass Rate
2007.1	32	5	27	84%	11	5	6	55%
2007.2	44	5	39	89%	9	3	6	67%
2008.1	38	3	35	92%	14	3	11	79%
2008.2	52	13	39	75%	11	4	7	64%
2009.1	43	3	40	93%	12	5	7	58%
2009.2	34	2	32	94%	9	5	4	44%
2010.1	45	12	33	73%	7	3	4	57%
2010.2	50	4	46	92%	15	5	10	67%
2011.1	18	4	14	78%	19	5	14	74%
2011.2	16	6	10	63%	13	3	10	77%
2012.1	15	5	10	67%	14	4	10	71%
2012.2	27	9	18	67%	28	10	18	64%
2013.1	25	7	18	72%	20	2	18	90%
2013.2	19	11	8	42%	15	7	8	53%
2014.1	27	10	17	63%	28	11	17	61%
2014.2	25	7	18	72%	31	13	18	58%
2015.1	127	15	112	88%	80	39	41	51%
2015.2	118	30	88	75%	80	58	22	28%
2016.1	110	42	68	62%	86	74	12	14%
2016.2	130	39	91	70%	74	64	10	14%
2017.1A	88	10	78	89%	46	19	27	59%
2017.1B	33	11	22	67%	36	23	13	36%

6 Analysis of Key Issues Arising in the Review

Bias in Examinations

6.1 Both the complaint and submissions emphasised that the Fellowship OSCE lacks objectivity and permits examiner bias (whether conscious or unconscious) to affect the integrity of the examination process. Due to the perceived subjectivity of the examination, some submitters viewed the Fellowship examination as a competitive process rather than a qualifying process, as an OSCE is intended to be. This has evidently led to a perception amongst some submitters that the OSCE facilitates racial bias to discriminate against candidates, and that it is used to systematically remove IMGs from the College training program, and that the College, as a whole, is racially discriminatory.

Patent racial bias

6.2 It is apparent that the disparity in pass rates – in conjunction with the limited information made available by the College about the operation of the new examination format and limited feedback given to candidates – has fostered perceptions amongst some trainees that the College's intent and actions in the transition to the new examination format was motivated by discrimination against NCCs. The EAG did not find any evidence of racism underpinning the College's rationale for the transition to the new OSCE format in 2015, nor in the College's conduct and management of examinations since.

6.3 However, it is concerning to the EAG that a small number of submissions did report individual experiences of race discrimination arising from interactions within the workplace and in the course of their training and assessments. This was also reflected in data provided to the EAG from the College's DBSH Survey, where 10.2% (n=186) of respondents surveyed reported they had experienced DBSH in ACEM assessments. ITAs and WBAs were nominated by over half the respondents as the areas where they had experienced discrimination, followed by over a quarter of respondents reporting that it had occurred in the OSCE. Of the respondents who reported DBSH in assessments, 30% of respondents had been in the training program for 6 - 10 years and 80% for >10 years.

Unconscious bias

6.4 Many submissions from trainees highlighted a perception that examiners pass or fail candidates by feel rather than by an objective assessment of capabilities and clinical skills. Submitters infer that, by facilitating a non-objective examination, the College is actively permitting racism through the disparity in outcomes experienced by NCCs.

6.5 Some submissions queried whether examiners were aware of natural biases and how this can affect their assessments. It was acknowledged in some submissions that actors and confederates are not immune to their own biases, which can create inconsistencies in the way in which they interact with candidates, thus impacting on candidates' responses.

6.6 Submissions from examiners offered counter views as evidence of why unconscious bias was not a principal factor in the disparity of pass rates: examiners are subject to orientation and training; are experienced educators who are aware of the risks of unconscious bias in judgement marking; there has been no significant change in the composition of the Court of Examiners, methods of examiner training or Examiner Code of Conduct since the introduction of the OSCE yet the pass rates of IMG candidates have fallen markedly during this time; and even if unconscious bias was present in examiners, it would need to be widespread to significantly affect the overall pass rate of candidates as each examiner contributes only 1 - 2% of the total mark of any one candidate.

Impact of stereotypes

6.7 Many stereotypes were reported in submissions about IMGs. Submitters indicated that these stereotypes were reported to them or witnessed by fellow trainees, FACEMs, and in some case DEMENTs and examiners. These included the following stereotypes about IMGs:

- 6.7.1 *IMGs' learning style* – That IMGs are not fast on their feet; are rote learners; have inflated expectations of their capability; and many who fail examinations tend to be reflex learners rather than problem solvers.
 - 6.7.2 *Preconceptions about standard of training and knowledge* – That some IMGs have less training than others, or training from a country with non-comparable health care standards, and, as such, are a liability to the health system; IMG candidates need to outperform their locally trained colleagues in order to pass, as the standard applied to the assessment of their performance is higher; IMG candidates from English speaking countries such as the UK are preferred by hospitals and are more likely to succeed at OSCEs.
 - 6.7.3 *Perceptions relating to cross-cultural competency* – Australian candidates are more direct and confident in their communications, and where other cultures are not, they are perceived as not capable of performing at the requisite standards; perceived inability of some cultures to demonstrate empathy in a way that will resonate with local patients; inferences that some IMGs of particular ethnic or religious backgrounds are likely to hold discriminatory attitudes towards others and treat them unfavourably.
- 6.8 Such negative and uncontested stereotypes and perceptions can lead to confirmation bias and impact on the way an individual's performance and ability is perceived and thus assessed. Confirmation bias, as the term is typically used in psychological literature, connotes the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand, often leading to the inappropriate bolstering of hypotheses or beliefs whose truth is in question.¹⁹

Bias and the Correlation with Diversity

- 6.9 Examiner bias is a well-established risk in clinical examinations. Bias of clinical examiners against some types of candidate, based on characteristics such as gender, ethnicity, previously seen candidate performance²⁰ or first impressions,²¹ would represent a threat to the validity of an examination, since such biases are 'construct-irrelevant' characteristics.²²
- 6.10 To reduce the risk of discriminatory bias, it has long been advocated that examinations attempt to mimic the demographic characteristics of their candidature by those of their examiners.²³
- 6.11 International studies consistently recommend greater diversity in examiners to counter these effects, including:
 - 6.11.1 increased diversity of examiners for better representation of candidates coming through training programs;
 - 6.11.2 ensure diversity in training simulated patients, including the type of cases in the examination;²⁴
 - 6.11.3 apply a diverse group of examiners (e.g. male/female, Caucasian/non-Caucasian, local graduates/IMGs, older clinicians/younger clinicians) as fairly as possible across OSCE administrations, so that no candidate experiences a greater preponderance of examiner backgrounds; and
 - 6.11.4 ensure that all examiners and simulated patients have cultural diversity training and robust and regular calibration training.

19 Nickerson RS. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 1998, Vol. 2, 175-220, p.175.

20 Yeates P, et al. 'You're certainly relatively competent': Assessor bias due to recent experiences. *Medical Education* 2013;47(9):910-922.

21 Wood T. Exploring the role of first impressions in rater-based assessments. *Advances in Health Sciences Education* August 2014, Volume 19, Issue 3, 409-427.

22 Construct irrelevance is the extent to which test scores are influenced by factors (e.g. mode of presentation or response) that are irrelevant (not related) to the construct that the test is intended to measure. Adapted from: American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (1999) Standards for educational and psychological testing. Washington, DC: American Educational Research Association, p.173

23 Denney ML, Freeman A., Wakeford R. MRCGP CAS: are the examiners biased, favouring their own by sex, ethnicity, and degree source? *Br J Gen Pract* 2013 Nov;63(616):718-725.

24 Esmail A, Roberts C. Academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between November 2010 and November 2012: an examination of data. *The BMJ*, 2013;347.

- 6.12 Some submissions felt that the composition of examiners and confederates used in OSCEs was not representative of the diversity of candidates and patients presenting to Australian emergency departments.
- 6.13 Analysis of College data by Professors Farmer and Schuwirth of the 2016.2 OSCE, and information obtained by the EAG, indicates that:
- 6.13.1 for the 2016.2 OSCE examiner diversity for country of qualification (93% of examiners obtained their primary medical qualification from a Group A country, the mean age of all examiners was 51.5 years, with a gender ratio of approximately 1:2 of female to male examiners) was comparatively low to the diversity of the candidates attempting that examination (who were more represented by Group B countries);
 - 6.13.2 there was a lack of 'at standard' criteria setting for the domains of communication, leadership and teaching;
 - 6.13.3 calibration of marking criteria was sub-optimal given the continuous improvement measures the College has implemented since; and
 - 6.13.4 it was not clear if, during the station construction process, the stations were reviewed for any possible bias.
- 6.14 Where there is comparative low diversity between examiners and candidates and limited calibration setting there is a risk of subjectivity and culturally laden assessments of a standard, which may disadvantage a culturally diverse candidate group.²⁵
- 6.15 The College's global criterion for 'just at standard' being referable to practice in an Australian ED is entirely appropriate and justified, in that it is designed to ensure that trainees are safe to practise as specialists in emergency medicine in Australia and New Zealand. However, the EAG considers the above weaknesses in the examination may contribute to the risk of individual examiner bias.
- 6.16 Further it is not clear to the EAG if, during the station construction process, the recruitment and assignment of examiners for each OSCE and individual stations, and the examination station calibration, takes into account the cultural diversity of the trainees, and of patients who may present in the setting of an Australian and New Zealand ED.
- 6.17 On this basis the EAG considers that greater planning for, and awareness of cultural diversity, in the management of the College's examinations is needed to reduce the risk of racial bias and disadvantage to culturally diverse candidates.

Accounting for Unconscious Bias

- 6.18 While the psychometric analysis of the 2016.2 OSCE undertaken by Professor Schuwirth confirmed that there is no way to determine whether any form of bias occurred in individual stations (either against Group A or Group B), there was no indication of any form of bias large enough to statistically detect and account for the difference in pass rates.
- 6.19 As noted earlier in this report, Professor Farmer's literature review identified very few international studies that have investigated the potential for examiner bias specifically arising from the perceived country of training or ethnicity in high stakes examinations in relation to lower success rates of IMGs in comparison to local graduates. Most studies identified were quantitative with a focus on testing for bias using actual examination data in large-scale data sets. Professor Famer's report provided the following examples:
- 6.19.1 A large British comparative study involving over 2,000 paired independent examiners in a postgraduate clinical skills test conducted by McManus et al found that no examiners showed significant gender bias, and only a single examiner showed continued evidence consistent with ethnic bias. This examiner was non-white and appeared to be systematically awarding relatively higher marks to non-white candidates.

²⁵ Esmail A, Roberts C. Academic performance of ethnic minority candidates and discrimination in the MRCPG examinations between November 2010 and November 2012: an examination of data. *The BMJ*, 2013;347.

- 6.19.2 In the postgraduate study conducted by Dewhurst et al, a significant interaction between candidate and examiner ethnicity was found for performance on communication skills stations in a clinical examination (P = 0.011), but not on clinical skills stations in the same examinations (P = 0.176). Analysis of overall average marks showed no interaction between candidate gender and the number of assessments made by female examiners (P = 0.151). Potential examiner prejudice or bias was significant only in the cases where there were two non-white examiners examining a non-white candidate. Any simplistic explanation in terms of examiner prejudice can be excluded, as a systematic bias would also be expected to be evident in clinical skills stations as well. The effect was statistically significant in the communication stations, in cases where two non-white examiners meet a non-white candidate. Here the data suggested that the non-white candidate was given a higher score than when compared with candidates seeing white/white or white/non-white examiner pairs.
- 6.19.3 In the UK, a large study of a postgraduate clinical skills examination conducted by Denney et al attempted to determine whether examiners were biased, 'favouring their own' by gender, ethnicity and degree source. There were four main findings on univariate analyses shown in order of importance (effect on station score) (maximum marks per station is 9 marks), and effect size (%).
- (a) IMGs receive a higher mark from UKG examiners than from IMG examiners (mean difference = 0.32 marks, 3.6%) P<0.001.
 - (b) UKGs receive a higher mark from UKG examiners than from IMG examiners (mean difference = 0.24 marks, 2.7%) P<0.001.
 - (c) BME candidates receive a higher mark from a BME examiner than from a white examiner (mean difference = 0.20 marks, 2.2%) P<0.001.
 - (d) Male candidates receive a higher mark from a female examiner than from a male examiner (mean difference = 0.08 marks, 0.9%) P<0.001.
- 6.20 It is apparent from Professor Schuwirth's analysis and Professor Farmer's literature review that for unconscious bias to have a meaningful effect on candidates' outcomes in OSCEs and to be a significant contributor to the persistent differences in outcomes between Groups A and B, it would need to be widespread and systematic amongst examiners.
- 6.21 That is not to say it is not occurring at all in the ACEM Fellowship OSCE. The College's approach to standard setting and use of the Borderline Regression Method to determine the cut-score and pass mark required for each OSCE is appropriate and consistent with international use. However, the effectiveness of this approach relies upon robust standard setting for each domain/station through effective calibration of examiners to achieve a consistent and fair understanding of how to judge when a candidate's performance is 'just at standard' for that station.
- 6.22 The EAG's opinion is that the College's management of the OSCE to date potentially enables systemic discrimination to manifest in some elements of the examination.
- 6.23 Some domains are inherently subjective in nature (e.g. communication, leadership and management, scholarship and teaching) and where there is no defined criteria for assessing performance, and insufficient calibration of a station, a candidate's performance in these domains is typically assessed by reference to an individual examiner's experiences and opinions of what the requisite standard of performance is for a consultant. Given the demographics and experience of examiners, this is likely to be informed by reference to dominant cultural values and norms expected in the practice of Emergency Medicine in Australia and New Zealand.
- 6.24 These elements may have disadvantaged trainees who are not native English speakers and/or have obtained their primary medical degree from another country either culturally different to Australia or from a non-comparable health care system. However, the EAG has not been able to quantify the impact of these elements.

Transparency and Accountability of OSCEs

Transition to new format

- 6.25 Some trainees submitted that following the introduction of the OSCE there was little information made available on what it consisted of and how to prepare for it. For example, some trainees noted that there was no information made available prior to the OSCE on station breakdowns, sample examination scenarios or the required pass mark prior to undertaking examinations.
- 6.26 Various concessions have been made by personnel within the College that the changes to the Fellowship Examination format and assessment could have been better communicated to trainees to enable them to better understand the process, how to prepare and how they would be graded. It was, however, also noted that by virtue of being a criterion referenced examination, the pass mark cannot be known or established prior, as the Borderline Regression standard setting method uses data from candidate performance to determine the pass mark for the examination.

Feedback provided to unsuccessful candidates

- 6.27 The EAG finds that since the introduction of the OSCE, unsuccessful candidates have not been provided with adequate feedback to understand the basis for their marks and identify areas for future improvement and practice. Issues of concern identified throughout this review include:
- 6.27.1 inconsistency in capturing feedback amongst examiners;
 - 6.27.2 inadequacies in the feedback provided to unsuccessful candidates: not being specific in the critique in terms of explaining why a candidate's response was not competent/at standard, or what gave rise to them being marked as having committed a critical error; and
 - 6.27.3 following the introduction of the OSCE the College has acknowledged there was a drop off in the feedback provided to individual candidates and the content of the evaluation reports on the overall performance of each OSCE that are made available to candidates.
- 6.28 Relevantly, the College acknowledges this as an area requiring improvement and has since undertaken the following steps to improve the quality of feedback:
- 6.28.1 examiners have been trained in what is expected in terms of recording written feedback for candidates – this has included an emphasis on more specific feedback on what the candidate did not do to meet the standard required in a station;
 - 6.28.2 introduction of an additional matter on candidate scoring sheets for examiners to record an area of concern (where a candidate's performance involves an omission or commission that may harm a patient); and
 - 6.28.3 continuing to monitor the quality of feedback provided by examiners.
- 6.29 Many submitters commented that the lack of transparency in current examination marking, the inadequacy of examination feedback, the publicised comments of senior examiners and the lack of explanation from the College for disproportionate results, had contributed to a strong perception of discrimination.
- 6.30 The EAG considers that the inadequacy of feedback and lack of transparency around the conduct and performance of the OSCE contributes to a perception that the OSCE is an inherently discriminatory process. The EAG urges the College to counter this perception by publishing as much information about the conduct and performance of each OSCE (that is reasonably appropriate in order to protect the integrity and reuse of stations), so as to enable unsuccessful candidates to better understand their performance comparative to other candidates who undertake each examination. This, coupled with improvements to provide individual candidates with formative feedback on their OSCE performance, will contribute to a greater sense of transparency and greatly assist unsuccessful candidates and their DEMENTs to identify areas for future improvement.

Associated College Processes

WBAs and ITAs

- 6.31 Although beyond the terms of reference, issues in in-training assessments (in the form of ITAs and WBAs) were a recurrent theme in the submissions received by the EAG.
- 6.32 The College undertook a limited data comparison to determine any correlation between WBA and OSCE results. It was noted that broadly there is a correlation between some outcomes of the WBAs and OSCEs, in that similar performance is noted in some aspects across both assessments from a broad pool of candidates.
- 6.33 It is possible that inadequate supervision and training of candidates in the workplace has led to candidates who were not ready and/or not competent sitting the examination or continuing in the training program. This was admitted by the College in response to the interim report. The College advised that former ITAs conducted at the end of term appeared of questionable utility and ability to identify trainees who needed extra guidance and training in order to progress to a standard of practice at a specialist level. A review of assessments undertaken during training is underway under the auspices of the Council of Education.
- 6.34 It remains beyond the scope of the EAG's terms of reference for the EAG to make any specific findings about discrimination in WBAs, however the EAG notes the work of the DBSH Project and that a review of in-training assessments is being undertaken.

The College's complaints process

- 6.35 Concerns expressed in submissions were that the College's current complaints process lacks transparency and finality in the manner in which the College resolves complaints.
- 6.36 The EAG notes that while the policy states that complaints are dealt with on a confidential basis, there is no guarantee of anonymity and it is not clear to what extent any reporting occurs on the complaints received by the College.
- 6.37 The EAG considers that the College process must focus on early intervention, with the following features: immediate acknowledgement of a complaint; prompt responsiveness (including keeping the complainant up-to-date where the resolution is drawn-out); and a response proportionate to the seriousness and impact of the complaint (i.e. where the complaint is serious, the responding investigation should occur quickly and, where appropriate, be outsourced to an external investigator).
- 6.38 While the policy provides for active participation of the Board including updates from the CEO, the policy does not take into account what will occur if a member of the Board or the CEO is the focus of the complaint. Additionally, the requirement to refer matters to a Complaints Committee (established on an ad hoc basis), with a requirement that the complaint documentation be provided to the Committee 21 days prior to its meeting, highlights a flawed process which does not anticipate the serious nature of some complaints. The nature and extent of the Board's involvement in the complaint is also not clear from the policy.

Culture of the College

- 6.39 Submissions reported a perception that the culture of the College is not welcoming, generates retribution and targets those who speak out. The result is that the majority of the cohort of complainants have remained anonymous in an attempt to ensure there are no negative repercussions in relation to their training.
- 6.40 Submissions report that discrimination broadly occurs throughout the College and places of employment for College trainees and FACEMs. This includes indications from these submitters that they felt they were not well supported by the College when issues were raised about employment or College matters.
- 6.41 In part, this issue may be resolved by working with employers to ensure a collaborative process when matters are raised by trainees and members in an employment situation.

- 6.42 The EAG was struck by the number of trainees who were fearful of providing submissions except on an anonymous basis, for fear of retribution by colleagues, supervisors and the College more broadly.
- 6.43 The EAG noted a very low uptake of the Employee Assistance Program (EAP) the College instituted following the complaint. In the EAG's opinion, the low uptake appears to demonstrate a lack of trust in the College process and a perception that the EAP lacked integrity.
- 6.44 The EAG has also received a recommendation from submissions, that a position of IMG Trainee Coordinator be established to support non-Caucasian candidates and international medical graduates throughout their training. This is in response to concerns raised that trainees do not have anyone to advocate for them. The EAG however notes that FACEMs who made submissions commented that on a number of occasions they had personally attempted to raise issues of transparency and marking issues with the College without success, as an informal way of supporting trainees.
- 6.45 The College must rebuild the trust of its trainees and members to ensure that it establishes a culture that does not tolerate or facilitate discrimination. This requires the College to take proactive steps to deal with the noted inequalities represented throughout the complaint, submissions and more broadly.
- 6.46 The DBSH project and Harassment Prevalence Survey involves parallel areas of work that will address some of the concerns noted in this report. The EAG expects that broader reforms relating to the prevention of DBSH will be addressed through this parallel process. The EAG acknowledges that DBSH matters play an integral part in the culture of the College and acknowledges the importance of the work being undertaken in that project.

7 Findings

- 7.1 The EAG's conclusion is that the likely cause of the disparity in pass rates between Group A and Group B countries is multifactorial and cannot be attributed solely to systemic discrimination and unconscious bias.
- 7.2 While the psychometric analysis has demonstrated that examiner bias did not have a statistically identifiable effect on the 2016.2 OSCE, individual examiner bias could not be ruled out completely. The EAG has identified a number of factors that may have had an impact on the examination outcomes. As scoring of candidates' performance involves independent expert judgement, it is potentially open to subjectivity and conscious or unconscious bias. The absence of documented marking criteria for candidate performance for 'minimum level of competence displayed' for domains such as communication, professionalism, scholarship and teaching, and leadership and management can lead to greater individual subjectivity in the evaluation of these domains and interpretation of a candidate's performance as 'just at standard'.
- 7.3 Others factors may also have played a part: relatively low examiner diversity (e.g. in the 2016.2 OSCE 93% of examiners were from Group A countries of origin and 7% were from Group B countries of origin), sub-optimal discussion of cultural sensitivity in calibration of OSCE assessment criteria, and the fact that raters' cognition is predominantly influenced by their own experiences, values and interests. Taken together, all these factors may have resulted in judgements giving rise to differential and potentially unfair outcomes for Group B candidates.
- 7.4 Various other hypotheses for disparity in the pass rates were suggested through submissions and information presented to the EAG. Although it was beyond the scope of the EAG's terms of reference to test all of these or quantify the impacts of these, the EAG considers the following factors as likely contributors:
- 7.4.1 The manner in which the College managed the transition to the new Fellowship Examinations, in particular the introduction of the new OSCE format could have been better; concessions have been made that there was a lack of information and clarity of expectations and support for candidates preparing for and attempting the examination.
- 7.4.2 The absence of entry to training requirements with the College (in comparison to most other specialty colleges). Until recently (2014) the College permitted trainees to enrol with limited registration. From this it is unclear whether unsuccessful trainees have not been sufficiently prepared for the Fellowship Examinations.
- 7.4.3 Currently, trainees are permitted unlimited attempts at the College's examinations. Consistent with other colleges, the more attempts a candidate has, the less likely they are to pass the examination due to the impact of initial, secondary and tertiary failures on a trainee's personal and professional wellbeing. From 2018 trainees will be permitted a maximum of three attempts at the Fellowship Examination and should they not pass in those three attempts they will be subject to consideration for removal from the training program. This may also have contributed to some trainees attempting the examination before they are ready.
- 7.4.4 The uncoupling of the written and clinical examinations with the introduction of the OSCE, and the poor reliability of the 2015.1 Fellowship Written Examination, may have resulted in more trainees attempting the OSCE repeatedly in succession in the last few years.
- 7.4.5 The demands for doctors to fill positions in Emergency Departments over time, and competitiveness of Consultant positions resulting in workforce supply demands, may impact on the preparedness of the trainees, influencing them to attempt the OSCE prematurely.

- 7.4.6 Some submitters indicated that the current WBA process may be flawed due to the pressures of workforce resulting in assessors being reluctant to give poor WBA scores lest it jeopardise their staffing given the potential for poor performance in a WBA to trigger remediation, thereby essentially rubber stamping trainees who are not clinically competent. It is possible this reluctance to accurately score WBAs may result in candidates believing they are ready to attempt the OSCE in circumstances when they are not. The EAG notes however that the WBA does not filter candidates for the examination as there is no requirement for DEMTs/Supervisors to confirm trainees' preparedness to undertake the examination. The EAG considers that the role of WBAs must be looked at in the context of future continuous improvement undertaken by the College and their impact on the OSCE.
- 7.4.7 The College has conceded that inadequate supervision and training of candidates in the workplace has possibly led to candidates who were not ready and/or not competent attempting the examination or continuing in the training program. The College advised that former ITAs conducted at the end of term appeared of questionable utility and ability to identify trainees who needed extra guidance and training in order to progress to a standard of practice at a specialist level.
- 7.4.8 Training time limits apply, with Provisional Training to be completed within five years, and Advanced Training to be completed within 10 years, and all training and assessment requirements completed within 12 years from the time of enrolment as a trainee. Accordingly, candidates who are not otherwise ready to undertake examinations are nevertheless taking them to attempt to finish their training within the required period or else be subject to removal from training.
- 7.4.9 Since the introduction of the OSCE, unsuccessful candidates have not been provided with adequate feedback to enable them to understand the basis for their marks and identify areas for future improvement and practice.
- 7.4.10 Deficiencies in the examination process (such as lack of 'at standard' criteria setting for the domains of communication, leadership and management, and scholarship and teaching, and suboptimal calibration of marking criteria), coupled with comparative low examiner diversity can give rise to the risk of subjectivity and culturally laden assessments of a standard, which may disadvantage a culturally diverse candidate group.
- 7.4.11 That there is a true difference in performance based upon the source of a candidate's primary medical degree; that differences in their medical training can result in some candidates not being up to the clinical standard required due to non-comparable training methods and assessments.
- 7.5 On this basis, the EAG does not expect the introduction of its recommendations will correct or equalise the pass rates of these two cohorts, as some of the contributing factors, are beyond just the issue of racial bias in assessments and go to less tangible legacy issues for the College, and educational and social factors which contribute to examination performance.

8 Recommendations

- 8.1 The EAG's terms of reference requested that it provide recommendations to the College on immediate, medium and long-term reforms, including in relation to structural arrangements considered necessary to eliminate discrimination in College assessments (TOR 4).
- 8.2 Additionally, the EAG was requested to advise the College on appropriate professional development activities, including relevant training for examiners and supervisors of training felt necessary to address issues of discrimination in College examinations and related process (TOR 5).
- 8.3 The EAG first wishes to acknowledge the continuous improvement measures the College has undertaken since the initial receipt of the complaint. These include improvements to the OSCE process as well as the initiation of the DBSH Survey and the activities of the DBSH group.
- 8.4 The following recommendations are made by the EAG for the College to consider and implement.

Recommendations Dealing with Legacy Issues

- 8.5 Throughout the course of this review it has been suggested to the EAG that a number of legacy issues may have contributed to the disproportionate pass rates between IMGs and local medical graduates. These issues included no barriers to entry to the training program; allowance of entry on provisional medical registration; and questionable utility of ITAs to identify trainees in need of extra guidance and whether they were at standard.
- 8.6 To address the impact of these issues, the EAG recommends that the College consider:
 - 8.6.1 Reviewing the requirements and selection criteria (already underway) for entry into the FACEM Training Program.
 - 8.6.2 Consider what support or alternative options to an award of Fellowship could be made available to trainees coming to the end of their training term and who are unlikely to satisfactorily complete their training or demonstrate they are at the standard required to become a FACEM.

Recommendations Aimed at Continuous Improvement of the Examination Processes

- 8.7 Following the commencement of the EAG review, the College has undertaken a number of continuous improvement measures in its process of working towards resolution of issues raised in the original complaint.
- 8.8 These include general measures aimed at defining 'just at standard' candidates and tackling unconscious bias. Measures in relation to the Written Examination include improving questions used in the Fellowship Written Examination and improving marking processes for that examination. Measures in relation to the OSCE include assigning unique identifiers to examination candidates, utilisation of calibration processes before and throughout the examination to ensure standardisation, modifying the marking sheet for the examination, improving feedback processes for candidates, providing examiners with written feedback about their marking performance and video recording OSCE stations for auditing purposes.
- 8.9 In light of the measures undertaken to date, as set out above at 8.8, the following recommendations are made to assist the College in its process of continuous improvement.

Conduct of the examination

- 8.10 The standard required for a pass (i.e. 'just at standard') in every station should be clearly articulated prior to the examination being undertaken. This will require the marking criteria for each domain tested at each station discussed at the outset of calibration (and documented whether by check list or an agreed and documented standard) in order for examiners to consistently apply the standard to marking throughout the duration of the examination. This should extend to all domains, not just medical expertise, with priority given to standardising an agreed standard expected in the domains of communication, leadership and management, and scholarship and teaching.
- 8.11 Where possible, utilise/expand the use of multiple examiners to increase the number of observations to re-establish the integrity and validity of examinations with external stakeholders.
- 8.12 Continue to progress the work being undertaken in relation to recording the examination and explore how it can be best used to reasonably and practicably facilitate reviews of borderline fails, support feedback to unsuccessful candidates, and examination auditing and training purposes.
- 8.13 Implement a process whereby candidates can provide feedback on examiners, confederates and actors, noting that such a process should not be disclosed to examiners until marks are received and collated. Feedback sought should also elicit commentary and suggestions for improvement of the process overall.
- 8.14 Explore ways to develop a more diverse group of examiner and confederate backgrounds and apply these various examiner and confederate demographics (e.g. male/female, Caucasian/non-Caucasian, local graduates/IMGs, older clinicians/younger clinicians) as fairly as possible across OSCE administrations. This may necessitate a review of the terms of reference of the Court of Examiners and any associated policy(ies).
- 8.15 Ensure that all examiners and simulated patients/actors have cultural diversity training.
- 8.16 Ensure that all examiners and simulated patients/actors have robust and regular calibration.
- 8.17 Examiners should continue to receive regular feedback on their performance, including marking performance and feedback provided through candidate review forms.
- 8.18 In relation to scoring and adjustment of scoring in examinations, ensure that there is appropriate standard setting and ensure great transparency is utilised in publishing examination pass/fail statistics, including in relation to specific domains, publishing sample questions and answers, and providing examination candidates with an overview of the stations and domains to be tested, including spread of marks for relevant domains.

Support for examiners

- 8.19 Provide examiners with structured training and support on:
 - 8.19.1 cultural awareness and unconscious bias in examination marking;
 - 8.19.2 how to conduct calibration of 'just at standard' for assessing domains; and
 - 8.19.3 examiner performance.

Trainee preparation

- 8.20 Consider development and implementation or contracting of a training program to assist trainees who have experienced difficulties with the Fellowship Examinations to acquire experience in preparation for and performance in OSCEs.
- 8.21 Provide trainees with resources for training preparation in each domain.

Examination feedback

- 8.22 The objective of feedback should be to assist unsuccessful candidates to understand their specific areas of underperformance, why they were assessed as below standard and to identify areas for improvement.
- 8.23 Feedback provided should meet acceptable standards (consider UK Academy of Royal Medical Colleges – Guidance in Standards for Candidates Feedback Nov 2015).
- 8.24 Return to publishing all examination reports with details of station contents and commentary on candidate performance, to enable unsuccessful candidates to understand how they performed in comparison to other candidates attempting the same OSCE.
- 8.25 Consider releasing past examinations to current and prospective candidates with expected answers and mark breakdown in order to assist trainees in their preparation.

Review of results

- 8.26 It is evident that the integrity of the College's examination process has been called into question by the complaint. To assist in rebuilding the integrity of the process, the EAG recommends that:
 - 8.26.1 Beyond the standard reliability and psychometric assessments, the College consider periodical audits of examination results in terms of:
 - (a) comparison of station/examiner marks for Group A and Group B to monitor ongoing issues;
 - (b) the extent of correlation between candidate examination results against WBAs; and
 - (c) to assess both the quality of the examination process and the training program and various assessments.
 - 8.26.2 Consider implementation of automatic trigger for review of marks where a candidate is identified as a borderline fail.

Associated College processes

- 8.27 The EAG undertook a review of the College's policies and procedures to analyse broad compliance management processes and determine the policies and procedures already in place to assist with continuous improvement measures.
- 8.28 The EAG found that while the *Reconsideration, Review and Appeals Policy* is in place, it is not examination specific and, as such, does not clearly articulate what applicants can expect from a review. The EAG recommends the policy be reviewed and revised to ensure it clearly described the parameters of review for examination candidates, its limitations and that such an avenue cannot lead to a change in results.
- 8.29 The following recommendations are made with regard to the broader College processes:
 - 8.29.1 Increase trainee diversity and new FACEM participation in ACEM decision-making, including encouraging trainees who have not previously been involved to participate.
 - 8.29.2 Increase cultural awareness and cultural diversity within formal College positions and decision-making roles.
 - 8.29.3 Review governance mechanisms in the context of the current College examiners, including training and currency of clinical practice.

Complaints policy

- 8.30 As noted above at 8.27, the EAG undertook a comprehensive review of the College's policies and procedures. This necessarily included policies and procedures for making and resolving complaints. The following recommendations are made specifically in light of the existing complaints processes.
- 8.31 The College should review its *Complaints Policy* as it currently has limited application and utility – to ensure the process is transparent, adequately acknowledges potential outcomes and resolution processes, and accords with best practice.

- 8.32 The review should consider how the policy applies and can be utilised by complainants to achieve an outcome. It should also look at the transparency of the policy, including the possible outcomes of a complaint and the general complaints process.
- 8.33 The College should consider outsourcing handling of the whistleblowers framework to an independent third party, to ensure whistleblower protections are implemented and to reduce concerns about retribution.

In-Training Assessments and Workplace-Based Assessments

- 8.34 The FACEM Training Program IITAs were not within the scope of the EAG's review; however, it is clear that these assessments are a part of the complex fabric related to the disparity between Group A and Group B candidate outcomes. Views expressed to the EAG indicated concerns about the ITAs. Accordingly, the EAG directs the College to the following points for consideration:
 - 8.34.1 Consider utilising the WBA and ITA process as a method to determine preparedness to undertake the Fellowship examinations, including whether formal feedback meetings ought to be held with candidates to provide feedback on their preparedness.
 - 8.34.2 Give trainees the option of a support person when they receive feedback on WBAs.
 - 8.34.3 Ensure that outcomes are discussed with each trainee by their DEMENT to ensure feedback is provided.
 - 8.34.4 Review feedback mechanisms, particularly following WBAs and explicit feedback on areas for improvement.
 - 8.34.5 Consider ways to alleviate conflict between WBAs and workforce considerations including the need for greater transparency in these processes.

In-training supervision

- 8.35 As with the ITAs, in-training supervision was not within the scope of the EAG's review; however, is part of the complex fabric related to the disparity between CC and NCC outcomes. The EAG makes the following recommendations regarding these processes:
 - 8.35.1 Develop and implement a process to support Advanced trainees who are struggling in the program, which identifies areas of improvement and how improvement can be achieved.
 - 8.35.2 Ensure that trainee employment exit interviews (i.e. when moving from one hospital to another) are conducted by a person other than the trainee's DEMENT and/or a member(s) of the Court of Examiners.
 - 8.35.3 Implement a process whereby trainees can provide feedback on their supervisors and DEMENTs.
 - 8.35.4 Provide additional guidance and training for DEMENTs and examiners on cultural awareness and diversity.

Support for trainees

- 8.36 Introduce a trainee mentoring program to assist all FACEM Training Program trainees throughout their training.
- 8.37 Introduce an IMG trainee coordinator (either at the College or individual training hospitals) for a period of two years to provide trainees with assistance whilst other measures are being put in place to manage the cohort of trainees.
- 8.38 Encourage employers and support DEMENTs to provide individualised training and support for trainees who are deficient in particular areas identified through ITAs or through OSCEs.

Culture of the College

- 8.39 The EAG has received submissions that identify divergent views on the culture of the College including that trainees and other members have experienced discrimination throughout training programs. The College must ensure that it takes the lead to proactively prevent discrimination occurring within its training programs and within its broader College structure. The EAG recommends:
- 8.39.1 The College take greater leadership in denouncing and responding to discrimination where this occurs in the context of the FACEM Training Program.
 - 8.39.2 The College work with FACEMs to ensure there is an effective and consistent approach to addressing discrimination, bullying and sexual harassment and ensure that it provides a protective environment for those who make complaints.
 - 8.39.3 The College review its *Discrimination, Bullying & Sexual Harassment Policy* and *Procedures for Resolving Discrimination, Bullying and Sexual Harassment Complaints* to ensure that it adopts a zero tolerance approach to discrimination, bullying and harassment, and clearly articulates the College's role and approach to responding to these allegations and complaints by establishing, promoting, upholding and publishing clear and consistent standards.
 - 8.39.4 The College work with employers including training hospitals to develop partnerships to ensure these issues are addressed in the employment setting.
 - 8.39.5 The College consider implementing approaches to ensure diversity in representation on the Council of Education and Court of Examiners.
 - 8.39.6 The College create an implementation plan for the recommendations in this review to account to trainees and members on the progress it is making with regard to addressing the recommendations. The plan will assist the College in demonstrating its integrity with stakeholders and ensure it is accountable for its actions going forward.

① Annexure 1 - Terms of Reference



TERMS OF REFERENCE

EXPERT ADVISORY GROUP (EAG) ON DISCRIMINATION

1. BACKGROUND

Discrimination has no place in the College and its role in emergency medicine training and education.

The College recognises that discrimination can have a serious impact on those affected by it: it demeans the worth of individuals; it prevents our people from reaching their true potential; and it causes the loss of highly desirable talent from our profession. The College accepts its responsibility to eliminate discrimination in its processes. Recent media has suggested that aspects of the College's examinations processes may have discriminatory impacts and outcomes

The College acknowledges receipt of detailed submissions from some trainees in relation to the conduct of the most recent Fellowship Clinical Examinations (OSCE) ("2016 OSCE"), which contain detailed allegations of discrimination and bias. Those submissions deserve careful review and investigation, and an appropriate response to any actual or perceived discrimination or bias.

The College is therefore establishing an Expert Advisory Group (EAG) to assess and advise on these concerns. The EAG has a broad remit of assessing discrimination in relation to College assessments, as well as any other College activities, as it relates to the College, its Fellows, trainees and IMGs, with particular reference to the 2016 OSCE and associated outcomes. The EAG will advise the College on its role, policies and processes and advocacy in relation to discrimination, and will evaluate the complaint received in relation to the 2016.2 OSCE.

2. COMPOSITION

The EAG will be chaired by Dr Helen Szoke, formerly Victorian Equal Opportunity and Human Rights Commissioner and federal Race Discrimination Commissioner with the Australian Human Rights Commission.

The Deputy Chair will be Professor Ron Paterson, former New Zealand Health and Disability Complaints Commissioner and New Zealand Parliamentary Ombudsman.

Other members of the EAG include:

- Professor Kichu Nair, Professor of Medicine and Associate Dean of Continuing Medical Education at the University of Newcastle, and Chair of the Workplace Based Assessment Committee at the Australian Medical Council;
- two (2) ACEM trainees nominated jointly by the current trainee member of the ACEM Board, Dr Naveed Aziez and the Chair of the ACEM Trainee Committee, Dr Jessica Forbes;
- a recently qualified ACEM Fellow who qualified for election to Fellowship following the passing of a Fellowship Clinical Examination (OSCE) held since the beginning of 2015, and whose primary medical education was obtained in a country other than Australia, New Zealand, the United Kingdom, Canada or the USA; and
- two (2) members of the current ACEM Board, Dr Yusuf Nagree, the Chair of the ACEM Council of Advocacy, Policy and Partnerships (CAPP) and Dr Simon Judkins, President-Elect.

The EAG will be supported by senior staff of the College and externally appointed advisors in relation to specific matters where technical advice is required.

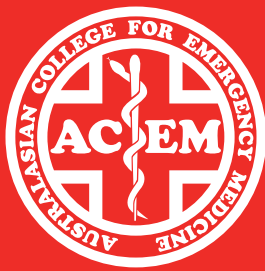
3. ROLE AND FUNCTIONS

The EAG will:

- 3.1. Consider the complaint submitted in relation to the 2016.2 Fellowship Clinical Examination and determine any remedy in relation to the complaint.
- 3.2. Undertake interviews and surveys, and review relevant literature, to understand the prevalence of discrimination in College assessments, including examinations.
- 3.3. Review College examinations (including policies, procedures and outcomes) to ensure they are not discriminatory and meet the expectations of internal and external stakeholders, including trainees, jurisdictions, and the public.
- 3.4. Recommend immediate, medium and long-term reforms, including in relation to structural arrangements, considered necessary to eliminate discrimination in College assessments.
- 3.5. Advise the College on appropriate professional development activities, including any relevant training for examiners and supervisors of training felt necessary to address issues of discrimination in College examinations and related processes.
- 3.6. Review the College's current initiatives and advocacy to prevent discrimination.
- 3.7. Ensure the College has an appropriate and robust complaints mechanism for discrimination, as well as a support program that is broadly available to Fellows and trainees, including IMGs undertaking the FACEM Training Program and SIMGs who have been assessed by the College as part of the processes conducted for the MBA and the MCNZ.
- 3.8. Review the College's Stakeholder and Communication strategies to respond to allegations of discrimination, to support trainees, Fellows and IMGs.
- 3.9. Establish a reporting framework where progress can be measured of the improvements in dealing with discrimination and in particular, in relation to College examination processes and procedures.

4. TIMEFRAME

To meet as required over a three month period and provide a final report / presentation to the College Board for consideration at its meeting scheduled for 19 June 2017, and, if more work is required, further reports as appropriate.



Australasian College for Emergency Medicine

34 Jeffcott Street
West Melbourne VIC 3003

t +61 3 9320 0444
f +61 3 9320 0400
e admin@acem.org.au

acem.org.au