

## **Theme: Assessment across the continuum/across borders**

### **Stream 1: Medical ethics in a cross-cultural setting**

- **Dr Roger Worthington, NHS Global Health Exchange**
- **Corresponding authors: \*Worthington R, \*Madhok R**

By its nature, ethics is influenced by local culture and prevailing legal frameworks, so developing a resource suitable for self-study that can be used by learners in developing countries presents various challenges. If materials are too generic, they lack focus and have no practical relevance; if they are too specific, they will not cross national boundaries, let alone continents. The NHS Global Health Exchange [GHE] [1] embarked on a project in 2015, in association with Peoples-uni, [2] to develop a course that could help fill a vacuum that exists in countries where practitioners and institutions have limited access to specialist staff and resources, and no formal curriculum for teaching medical ethics, such as in India and Uganda. The authors (with contributions from an international team) collaborated to produce a new, online course, launched in April, 2016, incorporating self-assessment, reflection and feedback. The course aims to be culturally sensitive, departing from widely-used frameworks, such as those originating in the USA in the 1970s. Content is designed to have practical relevance to contemporary learners, such that it could potentially influence health care on the ground, where, for example, significant problems exist in relation to professional standards [3]. By October 2016, the course will have been live for six months, which is a suitable point to evaluate user feedback and reflect on possible developments for the future as part of an iterative process.

### **Stream 1: An evaluation of medical student's knowledge of paediatric vaccinations in United Kingdom (UK)**

- **Dr Nighat Nadeem, King's College**

#### **Aims**

Medical Students now have decreased exposure to Vaccine Preventable Diseases (VPDs) as successful vaccination programs have decreased their prevalence. This combined with the media's negative portrayal of vaccines may cause misconceptions and misinformation. The aim of this study was to explore medical students' knowledge of paediatric vaccinations, highlight knowledge gaps, identify training needs and make recommendations for future training.

#### **Methods**

Vaccination knowledge of medical students from three UK medical schools was assessed by an anonymous, self-administered, cross-sectional, internet-based survey from 14 April 2015 to 14 July 2015. Questions addressed vaccine guidelines, schedules, administration, handling, contraindications and adverse events. Analysis included comparison of proportions with the use of descriptive statistics. Ethical approval was obtained from King's College, University of London.

#### **Results**

A total of 109 medical students participated from 3 London medical schools. Overall, 33/109 (30.3%) students expressed feeling 'not much' or 'not at all confident' with children's vaccines. The mean knowledge score of all students was 5.7/10 (57%). The most correctly answered question (answered correctly by 102/103 (99.0%) students) across all three medical schools was related to whether vaccines cause autism. The most poorly answered question overall was related to whether vaccines can be frozen to maintain their potency and was answered correctly by 31/103 (30.1%). The number

of students confessing to not knowing an answer rather than attempting to guess the answer was also highest for this question: 43/103 (41.7%).

## **Conclusion**

This study identifies gaps in knowledge amongst medical students in UK and the findings form a platform upon which to develop educational interventions which can be integrated into formal educational curriculum. Recommendations include developing up-to-date core competencies and promoting specific communication skills training in the role-play setting. Teaching methods used in various institutions should be analysed and compared to determine the most effective teaching strategies.

## **Stream 1: Can A Formative Integrated Clinical Anatomy Spot Test Predict Students' Performance in Summative Applied Medical Knowledge Assessments?**

- **Dr Tudor Chinnah, University of Exeter Medical School**
- **Corresponding authors: \*Chinnah T, Leitner C, Brandom K, Rice N, Devaraj V.**

Integrating knowledge of anatomy with other biomedical sciences is very important in clinical practice. We developed a formative integrated Clinical Anatomy Spot Test (iCAST), which focuses on students' ability to demonstrate clinically-relevant factual knowledge. This study evaluates its validity and predictive value on students' performance in summative applied medical knowledge (AMK) assessments. The iCAST was first piloted with 36 year 2 students in 2012/3 and rolled out to year 1 students in 2013/4 (n=130) and 2014/5 (n=129). It is delivered termly to years 1 and 2 students and requires students to rotate through 50 stations of integrated short-answer questions. Immediate feedback is provided, scripts are peer-marked and scores are independently verified. In both cohorts, average scores in the iCAST correlated with scores in the summative end of year 1 content-specific knowledge assessment (2013/4:  $r=0.46$ , p

## **Stream 1: A methodology to check the validity of OSCE cut scores derived by modified Angoff method**

- **Dr Steve Capey, Swansea University**

### **Introduction**

The setting of standards for individual OSCE stations is a controversial area; the debate between absolute standards and relative standard methods has by no way been resolved. We currently use a modified Angoff method to set standards on individual stations and make progression decisions based on the number of stations passed.

### **Methods**

We have collected data on our OSCE assessments that includes objective checklist data that is standard set by objective criteria. We also require our examiners to make an overall global judgement about the performance of the candidate on the station. We have analysed this data to evaluate the total number of fails by the Angoff method and the overall judgement by the examiners.

### **Results**

We found that the overall judgement of the examiners on the candidate correlates well with our Angoff scores on most stations. We have tracked whether the same candidates that fail by Angoff are the same as the overall scores.

### **Conclusions**

The checking of individual OSCE station cut scores with the overall number of failure and borderline students correlates well. Some differences remain with the individual candidates; however this may be related to examiner behaviour.

Take-home message: It is possible to estimate and check the validity of an Angoff cut score for an individual OSCE station by utilising global scores provided by examiners; allowing a useful post assessment check of the standard setting process.

## **Theme: Predictive validity and Patient & Public Involvement**

### **Stream 1: The predictive validity of the BioMedical Admissions Test (BMAT) for Multiple Mini Interview (MMI) performance**

- **Dr Sarah McElwee, Admissions Testing Service, Cambridge Assessment**
- **Corresponding authors: Cheung K, \*McElwee S, Shannon M**

#### **Introduction**

BMAT is used by medical and dental schools in their selection processes. It measures potential for demanding, science-based study by assessing across three sections: aptitude and skills (Section 1), application of scientific knowledge (Section 2), and written communication (Section 3). Previously, predictive validity studies have primarily used on-course academic performance as criterion variables; these studies have shown that Section 2 and Section 1 predict pre-clinical examination performance. However, admissions tests are commonly used with personal statements to shortlist candidates for interview, rather than to make final selection decisions onto a course, making it important to establish whether they predict performance at interview.

#### **Method**

Data on BMAT, personal statement and multiple mini interview (MMI) scores were collected for a single admissions cycle of a UK dental school, where shortlisting was conducted according to the school's pre-existing criteria. For 278 shortlisted applicants, scores on BMAT were correlated with personal statement scores and performance on a 9 station MMI.

#### **Results**

BMAT Section 1 and Section 3 scores correlated positively with overall MMI scores. At the level of performance on individual MMI stations, three correlated positively with Section 1 and two other stations had positive relationships with Section 3. There was no evidence of a positive relationship between scores assigned to personal statements and overall MMI performance.

#### **Discussion**

BMAT scores show small but significant correlations with aspects of MMI performance, whereas personal statement scores did not demonstrate similar suitability as a screening tool for interview. These results are discussed alongside established findings that, of the three BMAT sections, Section 2 (application of scientific knowledge) most strongly predicts on-course academic performance. An overview of BMAT predictive validity and the test's relationship with other selection tools, particularly personal statements, is presented.

### **Stream 1: Measuring knowledge growth: reliable change and identifying struggling students**

- **Dr Daniel Zahra, Plymouth University Schools of Medicine and Dentistry**
- **Corresponding authors: \*Zahra D, Cockerill J, Gabe-Thomas E, Roberts M.**

Within medical education it is important both to assess student knowledge and help those who do not attain the required level of knowledge; but how these two goals are achieved and how they relate to each other varies widely across medical schools.

Students enrolled on the BMBS programme at Peninsula Schools of Medicine and Dentistry undertake four single-best-answer multiple-choice format medical knowledge assessments in each academic year. These assessments are designed as progress tests, developed on the principles of frequent-look and rapid-remediation, and track changes in medical knowledge over time.

At present remediation decisions are based on a test-by-test basis, with students failing to achieve satisfactory scores being referred for remedial support. Basing such decisions on individual test performance might be considered at odds with the principles of progress testing, especially where performance criteria are norm-referenced.

The Reliable Change Index (RCI; e.g. Jacobson & Truax, 1991; Zahra et al, 2016) provides a way to evaluate the consistency of changes in student performance across multiple tests over and above the variability inherent in the assessment format. This shifts the focus of the analysis from conventional group-level statistics to an analysis based on the individual's performance. The current work investigates the possibility of using the RCI to complement existing remediation processes, in particular its usefulness in pre-emptively identifying students who may benefit from support and providing individually tailored feedback for entire cohorts.

### **Stream 1: Is a Secondary Task Method Appropriate for Measuring Mental Workload in Medical Students?**

- Miss Bryony Woods, Cardiff University
- Corresponding authors: \*Woods, B. Byrne, A.

#### **Context**

Mental workload (MW) is an abstract concept that sees cognition as a small and finite capacity to process conscious, logical thoughts. A secondary task (ST), an additional task added on top of the primary task, is one way of measuring MW. When workload of the primary task approaches capacity, ST performance will decrease giving an objective measure of MW.

#### **Objectives**

This study aims to validate the ST method as a measure of MW in medical students. It is expected that the measured workload will increase with task complexity.

#### **Methods**

Medical students from year 2 to year 5 at Cardiff University were recruited. The ST involved tapping the screen of an iPhone 5 when it vibrated at random intervals. The time taken to do this was recorded. Each participant completed four standardised tasks for a total of four minutes each, alongside carrying out the ST. Task 1 measured participants' baseline workload. Task 2 involved listening to a recorded history. Task 3 was undertaking venepuncture on a simulated arm and task 4 involved simulated venepuncture, alongside listening to another history.

#### **Results**

40 students were recruited. Measured workload increased with task complexity, p

### **Stream 1: Mentoring support for undergraduate medical assessments: Two-year evaluation of a near-peer mentoring initiative**

- Mr Timo Tolppa, Peninsula College of Medicine and Dentistry

- **Corresponding authors: \*Tolppa T, Glanville L.**

## **Background**

Peer mentoring has been shown to be a significant source of academic and professional support for those in the medical profession.[1-2] The additional support mentoring offers is welcome, especially during stressful transitioning times and throughout medical education to cope with new assessments.[3-5] We therefore developed a near-peer mentoring initiative across all years of three undergraduate medical programs in the South-West to assess the need for mentoring at various stages of training and to explore the potential benefits of mentoring on the assessment performance of both mentees and mentors.

## **Overview**

Following a successful pilot in 2013-14, 228 student volunteers were trained to mentor 436 mentees at the start of 2014-15 across all years and main localities (Exeter, Plymouth and Truro). During this academic year the initiative has been expanded to one additional site (Torbay) with over 180 mentors working together to mentor around 130 students. The initiative has been evaluated each year with end of year questionnaires and focus groups of both mentees and mentors. The focus groups at the end of this year will specifically explore the impact of the initiative on assessments.

## **Results**

The results from the pilot and first year of the initiative demonstrated that 87% of mentors and 57% of mentees significantly enjoyed participating. Assessments and academic support have been identified by mentees as the main benefits of the initiative. The existing results will be combined with the end of 2015-16 evaluation and presented with a particular focus on the benefits of near-peer mentoring on assessment performance.

## **Discussion**

Similarly to other mentoring programmes, the initiative has been well received by both mentors and mentees, with significant benefits to both.[6-9] The full results of the evaluation aim to further characterise the benefits of mentoring on assessment performance and the reasons behind this.

## **Theme: Preparedness for practice**

### **Stream 1: Use of peer feedback to enhance medical students' reflective writing**

- **Dr Rose Crowley, Whittington Hospital**
- **Corresponding authors: \*Crowley R, Fertleman C, Begley R, Winyard P**

## **Background**

Doctors are expected to provide constructive feedback to their colleagues and to reflect on their own practice. We sought to improve medical students' preparedness for these aspects of professional practice through an innovative peer feedback strategy for reflective writing. Reflective writing is a well-established strategy in medical education to promote deep learning and encourage self-awareness and well-conducted feedback can enhance this. Whilst peer feedback on reflection is relatively well established in teacher training, it has not been significantly evaluated for health professionals.

## **Objectives**

To establish whether medical students gain any additional learning from peer feedback rather than faculty feedback on reflective writing. To determine the benefits and challenges when students peer mark reflective work and give written or verbal feedback.

## **Methodology**

Seventeen UCL iBSc Paediatrics students anonymously marked a peer's reflective writing assignment using the same mark scheme as faculty markers. Students completed questionnaires on how useful they found the experience of reading another student's work, how valuable they found the marks and comments from their peer and faculty assessors and their experience of offering verbal feedback in a reflective group.

## **Results**

Many students highlighted benefits from peer marking, particular that it made them re-evaluate their own submission and rethink what constitutes useful reflective writing. There was no significant difference between peer and faculty marks and students highlighted benefits from each type of feedback. However, some students struggled to assign a grade to a peer's work, feeling they lacked the authority to assign even a formative mark.

## **Conclusion**

Students gained from assessing fellow students' work and from receiving peer feedback and this technique is now used as standard for their formative reflective writing coursework. The impact of this programme on their future reflective writing and the relative utility and acceptability of written and verbal peer feedback merit further exploration.

### **Stream 1: Can assessment information be used to understand how to learn a skill?**

- **Dr Rachel Davies, University of St Andrews**
- **Corresponding authors: \*Davies RC, Ford M, Tabberer IDM, Struthers JE.**

Medical training is critically dependent on the acquisition of skills, and much time is spent assessing these skills, however not much is known about how to help students acquire skills, short of saying 'go away and practice', and even that assumes that practising helps. The importance of reflection and feedback in undergraduate medical education is reinforced by the General Medical Council. Whilst the use of video and reflection in communication skills is well documented, its role in clinical skills development is less well established. The Medical School have developed a process of student self and peer appraisal of videoed clinical skills and personal reflection on performance. This research aims to evaluate whether engagement with this process (and thus enforced practicing) is linked to clinical skills OSCE performance.

We are collecting data from 1st year medical students at the University of St Andrews regarding engagement with the set clinical skills video tasks and aim to link this with how they performed in these skills in the OSCE. The OSCE includes stations which have had associated video and reflection tasks post teaching, and some that have not. For stations with video tasks, we propose to link performance to measures of compliance with the task set. We propose to evaluate how students who have met markers of engagement with the video tasks perform in comparison to their peers. We also seek to evaluate whether there is a significant difference in OSCE performance between skills associated with student video portfolio tasks, and those without.

### **Stream 1: Promoting cultural competence: Peer and tutor assessment of UK and Ethiopian students' reflections of ethical case scenarios**

- **Ms Annie Wood, Peninsula College of Medicine and Dentistry**
- **Corresponding authors: \*Wood A, \*Tolppa T, Sharp O, King J.**

## **Background**

Developing cultural competence is increasingly important in the era of globalisation, as doctors must be able to respond to the specific needs of patients from diverse backgrounds.[1-2] The GMC requires doctors to be culturally competent, yet contact with different cultures and assessment of cultural competence during medical education is limited.[3-5] With an aim to provide medical students exposure to different cultural professional values and assess their cultural competence, we created an ethical case scenario exchange programme between Peninsula Medical School and Wollega University Medical School in Ethiopia.

### **Method**

Small groups of students were identified and asked to write descriptions and reflections of clinical scenarios highlighting ethical dilemmas. Case folders were exchanged and students at both schools participated in facilitated discussions to explore different aspects of each scenario. Students provided peer feedback on the case reflections and UK-based facilitators commented on the quality of the scenarios. The use of these peer and facilitator-assessed reflective case portfolios to promote cultural competence has been evaluated using student questionnaires.

### **Results**

Two exchanges have taken place. The initial results revealed that students have gained a new global and cultural perspective of medicine, which was highlighted by similarities and differences in practices between the two student cohorts. Peer feedback from the UK highlighted that the cases chosen by Wollega students lacked ethical dilemmas and appropriate reflection. The full results of both the exchanges, including the facilitator feedback on the quality of case reflections, will be presented.

### **Discussion and Conclusion**

Overall, the programme has been well received by students. Ethics does not feature prominently in the Wollega curriculum, which may explain the students' limited analysis. Further peer feedback and exposure to scenarios will develop their understanding and reflections. The facilitator assessments will provide additional feedback on the reflective case portfolios and supplement student learning.

## **Theme: Professionalism including diversity**

### **Stream 1: Longitudinal Assessment of the Communication Skills OSCE: Implications for Performance Development and Student Diversity**

- **Mr. Chris Edwards, Cardiff University**
- **Corresponding authors: Edwards, C. Tayyaba, S.**

### **Background**

Communication skills (CS) are fundamental to the development of a good patient-doctor relationship and the GMC have made the inclusion of communication skills training in undergraduate medical curricula a necessity. Few studies - with conflicting results - have been reported so far regarding the development of CS.

The current study aims to investigate if there are any:

- Changes in student CS performance, as they progress through their undergraduate degree;
- Differences in CS and perceptions of the CS OSCE fairness, quality, and process across protected characteristics.

### **Methods**

This study longitudinally followed a cohort of Cardiff University undergraduate MBChB students (237), who graduate in 2016, as they progressed through clinical years 3, 4 and 5. CS performance was derived from the domain-based checklist mark sheet at the end of year OSCE examinations. The scores from CS content and process domains were combined to create an overall domain score. A 16 item survey was also created to investigate student perceptions into key OSCE domains.

## **Results**

A repeated measures ANOVA revealed significant differences between CS performances over the 3 examination years. A 2-way ANOVA revealed an overall significant effect of ethnicity on CS performance. Gender alone and gender alongside ethnicity had no significant effect on CS. A one-way ANOVA revealed that there were no significant differences in students' perception of the OSCE across ethnic minority and white students, and the majority of the students found the OSCE to be fair and objective.

## **Conclusion**

The study demonstrates that CS dip in the 4th year and rise in the 5th year examinations to a level not significantly different to the 3rd year. Ethnic students underperformed in CS compared to white students. Ethnicity had no significant effect on perceptions of the OSCE. Gender alone and combined with ethnicity had no significant effect on CS.

## **Stream 1: Development and Evolution of Progress Testing in Undergraduate Dental Education**

- **KAMRAN ALI, Plymouth University**
- **Corresponding authors: \*ALI K, ZAHRA D, SALIH V, TREDWIN C**

## **Objective**

To evaluate the Use of Progress Testing for Undergraduate Dental Students at Peninsula Dental School, UK

## **Methods**

Data were collected for Progress Tests conducted from 2007-2016 involving ten cohorts of Bachelor of Dental Surgery (BDS) students in seventeen sittings. The data were analysed in SPSS version 22.0 (SPSS Inc., Chicago, IL, USA) and R to identify the differences in the performance of students in successive years. Internal consistency of test scores was calculated using Cronbach's Alpha. Test-retest reliability of students' scores was assessed using the Pearson correlation along with 95 percent confidence intervals in successive sittings. Item facility was calculated for each of the 100 items in each of the progress test sitting for all cohorts. Analyses of variance (ANOVA) were used to identify variations in total scores, correct, incorrect and do not know responses based on demographic factors.

## **Results**

The test data showed satisfactory internal consistency as well as adequate test-retest reliability. The dental knowledge of students increases steadily over successive years as expected. The scores and correct responses mirror each other while, "don't know" responses decrease steadily. However, the incorrect responses stay relatively stable. Differences in the performance of students based on age, educational background, ethnicity and any known disability were also evaluated.

## **Conclusion**

Progress testing is a unique assessment tool which permits reliable longitudinal assessment of applied knowledge across the curriculum. There is merit in using progress testing for assessment of undergraduate dental student and monitoring the integration of applied dental knowledge during successive years.

## **Theme: Licensing exams, credentialing and revalidation**

### **Stream 1: The Evidence so far- a Guide to Assessment in Dental Education**

- **Julie Williams, School of Oral and Dental Sciences, University of Bristol University of Bristol**
- **Corresponding authors:**
  - o \*Julie Williams, School of Oral and Dental Sciences, University of Bristol
  - o Sarah Baillie, School of Veterinary Sciences, University of Bristol
  - o Sheena Warman, School of Veterinary Sciences, University of Bristol
  - o Susan Rhind, Royal (Dick) School of Veterinary Studies, University of Edinburgh
  - o Jonathan Sandy, School of Oral and Dental Sciences, University of Bristol
  - o Anthony Ireland, School of Oral and Dental Sciences, University of Bristol

#### **Aim of this work**

The term assessment derives from the Latin 'assidere' meaning to sit beside, suggesting that the assessor and the student travel together side by side on the journey to learn. Training the competent dentist requires evaluation against a series of standards. Our aim was to gather together current evidence for tools used for these two processes of assessment and evaluation as part of curriculum planning which, in turn, enhances the learning and development of successful dentists.

#### **Summary of work**

Methods used to assess the dental undergraduate and postgraduate were considered. A review of the literature included dental student, tutor and patient perceptions and evidence for the validity, reliability, educational impact, acceptability and cost of assessment methods.

#### **Summary of results**

A guide to dental assessment was developed based on the literature review and utilising a successful format already adopted in veterinary medicine<sup>1</sup>. The guidebook includes a short summary describing each assessment method and considerations for both new and experienced dental educators at the undergraduate and postgraduate levels.

#### **Discussion**

Synthesising the literature in an accessible format for colleagues aims to support staff development and on-going modernisation of assessment.

#### **Conclusion**

There is a body of evidence to support the use of a wide range of assessment methods although some score more highly in the utility equation than others.

#### **Take home messages**

This guide aims to promote the use of appropriate assessment methods within undergraduate and postgraduate dental education and is freely available online.

1. Baillie, S., Warman, S., Rhind, S. (2014). A Guide to Assessment in Veterinary Medical Education 2nd ed.
2. Williams JC, Baillie S, Rhind S, Warman S, Sandy J, Ireland A. (2016) A Guide to Assessment in Dental Education

Available from: [https://www.ole.bris.ac.uk/webapps/cmsmain/webui/\\_xy-7221180\\_1?action=ittach](https://www.ole.bris.ac.uk/webapps/cmsmain/webui/_xy-7221180_1?action=ittach). [Accessed on 23 March 2016]

## **Theme: Performance based assessments**

### **Stream 2: Developing a short OSCE to select a prize winning final year student**

- **Dr Elizabeth Metcalf, Cardiff University, School of Medicine**
- **Corresponding authors: \*Metcalf EP, Smith PS, Masefield R**

#### **Background**

Historically, final year medical student prizes at Cardiff University were awarded on the basis of a viva, assessing knowledge but not offering students a forum in which they could demonstrate their clinical excellence.

#### **Aim**

We sought to develop a clinical assessment that gave students the opportunity to demonstrate their skills and knowledge beyond the confines of a viva, and beyond a more structured, competency exam. We sought to create an authentic test, retaining patients at the heart of our assessment, as is the ethos of our curriculum.

#### **Method**

Since 2011, the top five scoring students at the end of their final year are invited to attend the 'Dean's' List' OSCE. The prize winner is selected using a five station OSCE- including medicine, surgery, women and children, general practice and laboratory stations. Students are given 20 minutes per station, during which time they perform a variety of tasks. Previous examples include assessing complex surgical and paediatric patients with multiple pathology, challenging communication simulations and integration of procedural skills with interpretation of laboratory results.

At the end of the OSCE, Examiners rank students on the basis of their clinical competence, reasoning and management, in addition to their knowledge. Following the assessment students meet examiners and patients in an informal, congratulatory setting.

#### **Results**

Anecdotally examiners and students alike enjoy the opportunity to stretch students in a more post graduate style exam than would be acceptable in a conventional OSCE. During 2015/16 we plan to conduct qualitative research to explore further the student's perspective of the assessment.

#### **Conclusion and recommendations**

The Dean's list OSCE rewards global competence. We are unaware of a similar OSCE in other institutions and therefore we would recommend similar strategies are adopted elsewhere in order to allow the most highly achieving students the opportunity to excel.

### **Stream 2: Students' evaluation of the OSCE experience in Cardiff Medical School**

- **Dr Saadia Tayyaba Centre for Medical Education, School of Medicine, Cardiff University, Cardiff Centre for Medical Education**
- **Corresponding authors: Tayyaba,S., Professor Hawthorne, K., and Professor Smith, P.E. Centre for Medical Education, School of Medicine, Cardiff University, Cardiff**

#### **Background**

Students as 'major consumer of the assessment' play a key role in determining the quality, acceptability, and usefulness of the clinical exams. Appropriate feedback from students could be used for improvement and refinement of the design, implementation, and teaching of clinical skills and competencies assessment. Unfortunately, research literature on students' evaluation of clinical assessment experience is still sparse and only a few studies have been reported to date in the published literature on students' perception of OSCE experience.

### **Objectives**

To evaluate MBBS final year students' perception in four key domains of OSCE experience (OSCE process, fairness, quality, and general attitude) and to investigate variation in these perceptions across diverse groups (gender, language, disability, and ethnicity).

### **Method**

Data were collected from 335 final years' students using a 27 items self-administered questionnaire given to students at the end of the OSCE. Ethical approval was obtained for the study. The sample included 8% students with reported disability, 68% females, 14% Welsh speakers, and 15% ethnic minority students.

### **Conclusion**

The findings clearly indicate a vast majority of the diverse groups of students (including over 80% home and ethnic minority students, students with reported disability, and language groups) had a higher overall agreement with the process, fairness, and quality of the OSCEs, and felt that the OSCEs accurately assessed their skills and competencies.

The overall perception on fairness of the OSCE was higher among diverse student groups and there was no significant variation in the perception of the OSCE fairness. Over 60% students agreed OSCE examiners were objective, not intimidating (62%), personality, gender, and ethnicity of the patient did not affect their performance (70%), they got enough information on what would be covered in the OSCE (78%). There were some comments that pointed out acute station as having insufficient time for scenario and skill examination.

## **Stream 2: Enhancing the quality of clinical assessments through improved examiner training and peer review**

- **Dr Elizabeth Metcalf, Cardiff University, School of Medicine**
- **Corresponding authors: \*Metcalf E, Goodfellow R, Masfield R**

### **Background**

There are strong internal and external drivers for a rigorous and transparent process of clinical assessment. Our underlying educational principle is to assess students upon authentic tasks to their future role as doctors. Students are expected to develop and demonstrate in increasing levels of complexity and integration their competencies of communication, procedural and cognitive skills in OSCEs and ISCEs.

### **Aim**

To ensure clinical assessments are reliable, reproducible and acceptable to stakeholders, it was identified that an enhanced approach to examiner training was required.

### **Method**

Mandatory pre-requisites for Examiners were defined:

- Examiner training < 2 years
- Equality & Diversity training < 3 years

- Current GMC (or GNC) Registration with a License to Practice
- Examiners MUST be aware of the standard expected from a student and be regular teachers of that year group

Face to face workshops supplement online training. They define the purpose of the assessments, discuss domain marking and standard setting followed by marking videoed stations and calibration of markers.

On the day of clinical assessments, parallel circuit examiners discuss the station guidance and content in order to further calibrate their marking and improve the reliability of the assessment before the assessment begins.

Peer review of examiners has also been introduced to:

- Refine training and development of examiners
- Supplement the external examiner process, ensuring quality and consistency, demonstrating transparency
- Improve standardisation and quality of the clinical assessments
- Provide examiners with constructive feedback to develop their education portfolios and share good practice

### **Results and conclusions**

Preliminary evaluation of these processes has found that Examiners find them both useful and constructive. Combining the previous psychometric examiner performance data with verbal and written feedback from peer observers will enhance the standardisation and quality of our clinical assessments. Further evaluation is planned over the coming year in order to fully assess the impact of these changes.

### **Stream 2: What is the best way to assess General Practitioner's fitness to practise?**

- **Keshwani K, University College London Medical School**
- **Corresponding authors: \*Sturrock A, Keshwani K, Mehdizadeh L, Valerio C.**

### **Background**

The GMC conducts Tests of Competence assessments for doctors referred for fitness to practise performance issues. In addition to knowledge test and OSCE, GPs also have a 'simulated surgery', devised to more closely reflect how they work. We sought the opinion of medical assessors conducting these assessments.

### **Summary of work**

45 total assessors were sent participation invite emails with a link to the online questionnaire, on a secure server with responses anonymised. Respondents were asked to express views regarding OSCE and SimSurg benefits. We asked respondents to rate agreement to statements about whether they assess the same competencies and which provides better feedback about candidates.

### **Summary of Results**

36 assessors responded. Inductive thematic analysis of free text responses revealed that assessors perceived OSCEs to be better suited to target specific knowledge and skills, as well as specific scenarios which SimSurg could not assess, such as colleague discussions. This targeted approach meant assessors perceived OSCEs to have lower fidelity compared to SimSurg, where scenarios more accurately portray a patient consultation. This authenticity was perceived to provide a more global picture of candidate performance. Specific skills, like prescribing, are assessed through integration within the consultation.

SimSurg is less intimidating as the assessor is not in the room, and does not penalise candidates who are not accustomed to OSCE format. Assessors felt SimSurg had poorer reliability, with greater risk of scenario going off-tangent. 70% of respondents agreed that OSCE elimination would still produce a valid assessment.

### **Discussion**

Three methods of GP assessment is stressful for candidates. Assessors felt that both methods were valid and had merits, but that SimSurg also gives an overall GP ability impression.

### **Conclusion**

A simulated surgery assesses many of the same domains as an OSCE, but with higher fidelity to the clinical practice of general practitioners.

## **Theme: Simulated patients/simulation**

### **Stream 2: Summative clinical assessment of medical student's ability to provide immediate care in a simulated medical emergency**

- **Dr Elizabeth Metcalf, Cardiff University**
- **Corresponding authors: \*Metcalf EP, Frost PJ, Goodfellow R, Kinnersley P, Williams S, Sillars A**

### **Background**

There is clear evidence that patient harm results from inadequate early clinical care, when patients present as an emergency<sup>1</sup>.

The GMC stipulate that graduates must be able to provide immediate care in medical emergencies<sup>2</sup>, recognise and assess the severity of a clinical presentation, formulate appropriate diagnoses and provide immediate life support.

### **Aim**

To develop a summative assessment tool that will ensure medical graduates from Cardiff University have demonstrated the necessary clinical skills to provide emergency care to their patients beyond graduation.

### **Method**

Since 2007/8, 'Final' summative clinical assessments in Cardiff have included the immediate management of acutely unwell patients.

The station format has evolved over time, as has the marking criteria and relative importance of acute skills assessment in the exam. In the early years, students were asked to make an 'ABCDE' assessment of a simulated patient, based upon a short clinical vignette.

Currently students must demonstrate their ability to manage both medical and surgical emergencies- integrating communication and technical skills, with clinical reasoning and care planning. Students take a brief history from a simulated patient, before reviewing clinical data, performing a procedural skill and answering standardised questions regarding differential diagnosis and initial management.

Simulated patient training is key to the effectiveness of these OSCE stations and is led by the clinical assessment lead in conjunction with critical care colleagues as necessary in order to ensure authenticity.

### **Results**

Reliability psychometrics are positive, as is feedback from external examiners.

### **Conclusion**

As a result of the evolving assessment, and in response to content expert recommendations, we have a novel, authentic, valid, reliable and feasible test of acute clinical skills.

### **Recommendations**

Further planned developments include research exploring cohort performance through an evolving curriculum and increased fidelity of simulation in a summative assessment setting.

## **Stream 2: Using teenagers from local schools in Paediatric History taking OSCE stations**

- **Dr Rachel Brooks, Cardiff University School of Medicine**
- **Corresponding authors: \*Brooks R, Pickerd N, Powell C, Metcalf E.**

### **Background**

The GMC state that 'effective communication between doctors and young people is essential to the provision of good care' (1). Cardiff medical students have paediatric specific communication skills taught before they go out on placement (2).

### **Aim**

To develop and evaluate an authentic and reliable OSCE station to assess the student ability to communicate with young people.

### **Method**

Teenagers from local comprehensive schools were recruited via their drama departments to become simulated patients. Scenarios were developed requiring the student to take a history of a common paediatric symptom and formulate a differential diagnosis. The information was put together in one side of A4 that was accessible to a teenage audience and divided into information that could be given spontaneously when asked about the presenting complaint and information only to be revealed if asked directly. The teenagers were trained in a session at school run by a paediatrician and an actor skilled as a simulated patient as close as possible to the date of the exam.

During the station the student is asked to take a focussed history from the teenager about the presenting symptom. Following this the student is asked to present their differential diagnosis, the diagnosis they think is most likely and why, and how they would proceed to confirm this.

Feedback was obtained from teenagers taking part and examiners were asked for their views on the simulated patients as part of their station feedback.

### **Results**

34 teenagers from 3 schools provided feedback after acting in one or two of five different scenarios and the experience had been positive for them. Many wanted to be invited back. Examiner feedback was positive. Reliability statistics were also positive.

### **Conclusion**

Links with local school has allowed us to develop authentic and reliable paediatric communication skills OSCE stations. School inspections and exam periods can be barriers otherwise schools are keen to be involved each year.

## **Stream 2: Simulation-Based Medical Education (SBME) in Otolaryngology surgery**

- **Dr Dione Lothar, East Surrey Hospital**

### **Introduction**

True airway emergencies are an uncommon clinical presentation in general medical practice. However, surgical doctors (especially those working within Otolaryngology departments) may be required to care for patients requiring invasive airway intervention, such as cricothyroidotomy. Furthermore, junior Otolaryngology trainees are required to provide emergency on-call airway services, sometimes without immediate senior supervision or support. Despite this, currently, no formal curriculum around airway management exists within the core surgical training program. SBME enables trainees to learn within a controlled environment, provides exposure to rare clinical scenarios (such as airway emergencies) and in addition, offers the opportunity for formative assessment and feedback without compromising patient safety.

### **Objective**

Design a SBME program utilising intermediate-fidelity simulation to teach junior Otolaryngology trainees the clinical and non-technical skills required to confidently manage airway emergencies and function safely as a member of the wider team.

### **Methods**

Three skills stations that provide the opportunity for hands-on practice in basic airway skills (simple adjuncts) and intubation, surgical airways (both cricothyroidotomy and tracheostomy) and fiberoptic nasoendoscopy. These stations will be followed by a series of simulation scenarios using a SimMan patient simulator; all with the common theme of management of patients with rapidly deteriorating airways.

### **Conclusion**

SBME is proving to be important in airway management education and as such may be an invaluable adjunct to both junior and higher Otolaryngology training. Furthermore, SBME in airway management may offer the only opportunity for formal teaching around this topic during core surgical training before junior trainees are faced with a real life patient with airway compromise.

## **Theme: Standard setting/psychometric analysis**

### **Stream 2: Comparison of Cohen and Angoff methods of standard setting: is Angoff worth it?**

- **Dr Laura Woodhouse, Newcastle University**
- **Corresponding authors: \*Woodhouse L, Lunn B and Kennedy DJ**

Standard setting is an essential process in defining competency in medical education. However, currently there is no gold standard method of standard setting. One commonly used method is Angoff in which a panel of experts estimate the percentage of borderline students predicted to correctly answer each question in an examination. This method is costly, time consuming and relies on the assumption that the panel can accurately define the borderline student. Recently the Cohen method has been developed and subsequently modified, to overcome these disadvantages. Our aim was to compare the standard set using our current method of Angoff to Cohen and modified Cohen

methods, to inform future standard setting practices. Cohen and modified Cohen methods were applied to historical data for written examinations across all 5 years of the MBBS programme at Newcastle University. Data included cohort sizes of 250-470 students per year, from academic year 2011/12 onwards. For single best answer (SBA) examinations in years 1 and 2, the Cohen method produced consistently higher pass marks and failure rates compared to Angoff. However, the modified Cohen method produced pass marks and failure rates comparable to Angoff. For year 3 SBA examinations, modified Cohen also produced comparable pass rates and failure rates to the Angoff. For years 4 and 5 SBA examinations the Cohen and modified Cohen methods produced comparable pass marks which were marginally higher than that determined by Angoff. This would have led to a 1-5% increase in the failure rate. With modified Cohen producing comparable standards to Angoff it suggests that this method may be a valid and economical alternative to standard set SBA examinations. We are currently analysing historical data for clinical examinations to determine whether standard setting using the Cohen or modified Cohen method is feasible for practical as well as SBA examinations.

## **Stream 2: MRCPsych Written Examinations – rationalising the content**

- **Mrs Kiran Grewal, Royal College of Psychiatrists**
- **Corresponding authors: \*Grewal, K., Wright, D., Husbands, A.**

Royal College of Psychiatrists' MRCPsych has traditionally consisted of 3 written examinations, each with 200 questions. A feasibility study of reducing a) the number of examination papers to 2, and b) the number of items in each paper, all whilst maintaining current reliability (0.88-0.95 in 2014), considered good/excellent for such examinations (George and Mallery, 2003), was undertaken.

Stage a) consisted of reviewing and remapping the syllabus from across 3 papers to 2, creating mock Papers, investigating their psychometric properties and conducting equality analysis of candidates' performance in them. Stage b) used the 200 item papers as references from which several random proportions of questions of the paper were selected and analysed for reliabilities. This anticipated reliability of such a shorter test was confirmed using the Spearman-Brown prophecy formula.

Analysis for stage a) showed no significant difference in paper reliability, pass rates, test scores, standard deviation of results or SEMs between current and new papers. Equality analysis found performances by most candidate groups remained the same, whilst the performance gap between PMQ/non PMQ candidates reduced.

Analysis for stage b) found that having only 50% of the items would give cronbachs of 0.83-0.88, and 60% gave 0.85-0.91.

In conclusion, the creation of 2 papers with less questions overall will not compromise desirable paper statistics, and within these, 120 questions yields the same quality of information as 200 questions. Streamlining the papers may have added benefits of reducing the performance gap between key groups.

Such a change would have financial, administrative and time resource benefits for the organisation, and time, cost and emotional investment benefits for examinees (Wainer and Feinburg, 2015).

Practical considerations include transitional arrangements, candidate feedback and quality assurances of items used.

Ultimately, smaller written examinations with good quality paper design and item selection yielding high quality information should be desired.

## **Stream 2: A Method for Ensuring Students' Medical Knowledge is Assessed in all Areas**

- **Ms Josephine Cockerill, Plymouth University**
- **Corresponding authors: \*Cockerill J, Ferguson C, Gabe-Thomas E, Zarha D.**

The BMBS programme at the Plymouth University Peninsula Schools of Medicine and Dentistry (PSMD) will be enrolling the fourth cohort of students later this year on to the five-year programme.

Each academic year students sit 500 single best answer medical knowledge questions across four tests “ up to a total of 2500 questions over the duration of the programme. Each question is categorised against an internal blueprint of 17 subject areas and 15 skill domains.

However, to ensure a balanced and comprehensive assessment of knowledge we need to make sure students are assessed across all of these skills and subjects. The current work describes how this is achieved by PSMD and provides an example for other medical schools facing similar challenges.

Our solution was to produce a blueprint heat map for each test to ensure a reasonable spread of questions, along with a cumulative heat map for each cohort to include all tests taken. We approached this by streamlining the data storage and by trialling various statistical packages to produce accessible heat maps displaying the number of items categorised under each possible combination of subject and skill domain prescribed by the PSMD blueprint.

The psychometric team now routinely produce these which are considered as part of the post-test review process. The cumulative heat maps contribute to identifying any subjects and domains with scant coverage needing to be included in future tests, or to be highlighted to the question writing panel if more questions fulfilling a blueprint criteria are required.

This new method of evaluating assessment coverage will ensure that as we move towards graduating our first cohort of students in July 2018 that their medical knowledge has been tested across all appropriate areas, as well as providing evidence should the coverage of medical knowledge testing be required by external sources.

### **Stream 2: Systematic evaluation of teaching qualities of surgical clinical teachers: psychometric properties of the modified SETQ tool**

- **Ahmed Al Ansari, MBBCh, MRCSI, MHPE, PhD, Consultant in Medical Education, Director of the Training and Education Department, Bahrain Defense Force Hospital. Assistant Professor of General surgery Arabian Gulf University, Senior Clinical Lecturers in medical education RCSI Bahrain**
- **Corresponding authors: \*Al Ansari A, \*Erekat Mona**

#### **Background**

Effective clinical teaching is crucially important for future patient care. Strong clinical training therefore is essential to produce strong physicians capable of delivering high quality health care. Tools used to evaluate medical faculty teaching qualities should be reliable and valid. This study investigates (i) the teaching qualities of clinical tutors, (ii) improving the teaching qualities of clinical tutors, (iii) modifying the System for Evaluation of Teaching Qualities (SETQ) instrument, and (iv) assessing the reliability and the validity of the modified SETQ instrument.

#### **Methods**

This cross sectional multicentre study was conducted in four teaching hospitals in the kingdom of Bahrain. 150 medical students were invited to evaluate 35 surgeons using the SETQ instrument between January 2015 and March 2015. Questionnaire feasibility was analysed using response rates, the average time required to complete the form, and the number of raters required to produce reliable results. Instrument reliability (stability) was assessed by calculating the Cronbach’s alpha coefficient for the total scale and for each sub scale (factor). To provide evidence of construct

validity, exploratory factor analysis was conducted to identify which items on the survey belonged together and were grouped as factors.

### **Results**

A total of 87 medical students completed 391 evaluations of 35 surgeons. The response rate was (58%) for the SETQ evaluation. Factor analysis showed that the data on the questionnaire decomposed into 6 factors that represented 71.7% of the total variance. Cronbach's alpha was 0.92 and higher for the six scales on the modified SETQ survey. The item-total correlation was above 0.40 for all items within their respective scales.

### **Conclusion**

Our modified SETQ questionnaire was found to be both reliable and valid, and was implemented successfully across various hospitals around the Kingdom of Bahrain.

## **Theme: Technology enabled assessment**

### **Stream 2: Very Short Answer Questions: A novel online assessment tool**

- **Dr Samantha Field, Imperial College London**
- **Corresponding authors: Sam AH\*, Field SM\*, Van der Vleuten C, Wass V, Schupke K, Harris J, Meeran K**

### **Background**

Single Best Answer (SBA) questions assess recognition rather than recall. Open-ended questions assess the ability to generate an answer and are considered more valid, but their use is limited by resource-intensive marking. We developed an online assessment system that could efficiently mark open-ended Very Short Answer (VSA) questions.

### **Method**

A 60-question formative examination was given to 299 medical students in SBA and VSA formats sequentially. The VSA questions were provided on a tablet with the same clinical scenario and lead-in as the SBA questions and a space to type a short answer. The VSA test was sat first by 155 students (VSA1/SBA2), whereas 144 sat the SBA version first (SBA1/VSA2). The results between the two cohorts were compared to assess reliability and validity. We evaluated the feasibility of VSA delivery and collected the students' opinions to assess potential impact on learning behaviour.

### **Results**

Two examiners reviewed the machine-marked VSA answers taking on average 1.36 minutes per question. Reliability was high: VSA1 (alpha=0.91) and SBA1 (alpha=0.84). The mean performance of the SBA questions in the two cohorts was similar (68.2% vs 69.7%, p=0.296). In the VSA1/SBA2 group, candidates scored significantly higher in the SBA2 (68.2%) versus VSA1 (52.4%)

### **Stream 2: A survey of current UK practice in videoing high stakes OSCEs in undergraduate and postgraduate medicine.**

- **Mrs Rhianna Hogley, Manchester Medical School**
- **Corresponding authors: \*Hogley H, Thampy H, Fisher J.**

The use of technology is rapidly altering the means by which many aspects of medical programmes are being delivered and OSCEs are no exception to this. At Manchester Medical School, there have

been repeated requests by students for the summative, high-stakes OSCEs to be videoed, in order that grades might be appealed and stations re-marked using video footage. This study was designed in response to student feedback in order to investigate the potential benefits, drawbacks and feasibility of videoing summative OSCEs. It includes a review of the literature on videoing OSCEs, a survey of the current practice of videoing high stakes OSCEs in UK medical schools and postgraduate medical examinations and semi-structured interviews. The semi-structured interviews are being used to explore the experiences and views of assessment leads in those medical schools and postgraduate medical institutes that are currently videoing high stakes OSCEs or that have seriously considered doing so.

At present, the research is ongoing, with 50 undergraduate and postgraduate medical institutions having been contacted. The response rate has been good, with 60% of questionnaires having been returned. 3 of the responding institutions do currently video high stakes OSCEs and their reasons for doing so are varied. 12 further institutions have seriously considered videoing high stakes OSCEs. While research is still ongoing, the findings so far indicate that the attitudes of assessment leads towards videoing for student appeals are wide-ranging and divided. In those institutions in which videoing has been seriously considered, many assessment leads express concerns about the potential costs of videoing and the technical challenges involved. However, the use of videoing for formative means, and for examiner training, are emerging themes cited as potential benefits.

## **Stream 2: Evaluating the Impact of Electronic OSCE Marking in Manchester Medical School**

- **Mr Michael Pollitt, University of Manchester**
- **Corresponding authors: \*Pollitt M, \*Thampy H**

Medical students increasingly want to receive high quality feedback from both formative settings in clinical placements and from their summative assessments. Common to many UK medical schools, Manchester Medical School (MMS) uses OSCEs across all five years of the programme to allow assessment of clinical competency. Traditionally the feedback provided by examiners to students had been handwritten. In 2015, MMS adopted an electronic means of OSCE marking using iPad devices. This significant change brought with it a number of clear logistical benefits including rapid collation and analysis of scores and typed legible, feedback immediately after each OSCE cycle and a reduced paper burden (especially important given that our OSCEs are simultaneously conducted across four sites). Furthermore, the transition to electronic marking has reduced error rates (i.e. those mark sheets which require a manual intervention) from around 25% to around 0.5% per OSCE. Full costing analysis revealed a saving of around 16k per annum.

In addition to these economic and practical benefits we wished to ascertain students' views of the pedagogical impact of electronic marking. We aimed to determine whether students who had experienced both feedback systems felt there were differences in the quality or quantity of the feedback they received (handwritten versus electronic) and any perceived impact on learning.

An electronic survey was sent to our final year MBChB students (n=431). Responses are still incoming and as of 15/06/2016 93 have replied (response rate of 21.58%). Emerging results suggest that respondents feel that electronic marking has improved both the quality and quantity of feedback when compared to handwritten. In addition students report that electronic feedback has improved their capacity to use OSCE feedback to trigger on-ward learning more so than with written. Detailed results will be presented.